Practice accreditation in general practice

Measurement validation and impact on chronic care management

Arna van Doorn - Klomberg

Practice accreditation in general practice

Measurement validation and impact on chronic care management

For reasons of consistency within this thesis, some terms have been standardized throughout the text. As a consequence the text may differ in this respect from the articles that have been published.

The studies presented in this thesis have been performed at the Scientific Institute for Quality of Healthcare (IQ healthcare). This institute is part of the Nijmegen Center of Evidence Based Practice (NCEBP), one of the approved research institutes of the Radboud University Medical Center.

The studies in this thesis were supported by the Dutch College of General Practitioners (Nederlands Huisartsen Genootschap), Utrecht.

Financial support by the Scientific Institute for Quality of Healthcare (IQ healthcare) for the publication of this thesis is gratefully acknowledged.

Nijmegen, 2015

Cover design: Sander Klomberg - van Doorn
Lay-out: Jolanda van Haren / Sander Klomberg - van Doorn / Arna van Doorn - Klomberg
Print: GVO drukkers en vormgevers B.V. | Ponsen & Looijen, Ede

ISBN: 978-94-6279-036-0

Practice accreditation in general practice

Measurement validation and impact on chronic care management

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen
op woensdag 3 juni 2015
om 10.30 uur precies

door

Arna Linda van Doorn - Klomberg

geboren op 30 oktober 1981
te Nijmegen

Promotor:	Prof. dr. M.J.P. Wensing
Copromotoren:	Dr. J.C.C. Braspenning Dr. R.J. Wolters
Manuscriptcommissie:	Prof. dr. J. van der Velden Prof. dr. P. Denig (UMCG) Dr. L.D. Elving

CONTENTS

1

General introduction

'Quality measurement can be compared with a hot spring. It is difficult to reach its core, but various aspects of quality can be observed like the mineral deposits around the spring.'

This thesis focuses on practice accreditation as a strategy to improve chronic care management in primary care. Practice accreditation is a structured method for assessing an organization. The introduction of this thesis starts with an outline of chronic care management and the use of audit and feedback to improve quality of care. Accreditation is a specific type of audit and feedback. The subsequent section elaborates on two areas of research in this thesis, which relate to different aspects in the evaluation of an accreditation program. The first concerns the performance measures used in practice accreditation and their metric properties. The second concerns the impacts of the Accreditation program on chronic care management and the associations of patient characteristics and organizational determinants with high quality of care. A total of 6 research questions are formulated which have guided the studies presented in this thesis.

CHRONIC CARE MANAGEMENT

In the Netherlands, as in other countries, chronic conditions place a large burden on the healthcare system[1]. Worldwide, the four leading causes of death are chronic conditions: cardiovascular disease, cancer, chronic respiratory disease and diabetes[2]. It is estimated that the prevalence of chronic conditions will increase in the next decades, mainly due to an ageing population and a rise in disease-specific risk factors such as obesity[3]. Furthermore, the number of people with multiple chronic conditions will also increase[4]. The main four chronic conditions that are commonly found in patients with high care costs include coronary artery disease, diabetes, congestive heart failure and chronic obstructive pulmonary disease (COPD). The prevalence of each of these diseases in the population is estimated between 1 and 10.1%, depending on the type of condition[5-8].

The organization of care for patients with a chronic condition differs across healthcare systems. The Chronic Care Model (CCM)[9] provides a structured framework to organize chronic care management. Several reviews have shown that interventions targeting (aspects of) this framework can improve chronic care[10,11]. For instance, Van Lieshout et al. compared practice characteristics of 11 mainly European countries using the CCM[12]. They found that stronger primary care was associated with better management of patients with established cardiovascular disease, but not in patients with high cardiovascular risk.

Since 2000, changes have taken place in the Dutch healthcare system that probably have affected several aspects of the organization of chronic care as described in the CCM. In 2006, the Dutch health insurance system changed into a more market-oriented health system. Although primary care was exempted from this, it was affected indirectly, for example through an increased workload due to substitution of hospital care, mainly regarding patients with chronic diseases. In the Netherlands, most patients receive treatment in primary care from a nurse practitioner (NP) and a family physician (FP)[3]. With the introduction of the nurse

practitioner in primary care, many general practices have restructured their chronic care management. Systematic reviews found that there are no differences between care provided by an NP (usually with physician backup) or by a physician alone, except that patients were more satisfied with care if the NP was involved[13,14]. Furthermore, interventions have taken place to improve chronic care regarding several other aspects mentioned in the CCM, including decision support[15] and improved registration.

Given the changes in chronic care management and the importance of this topic as measured by the burden on the health system, there is a need to evaluate the quality of care that general practices can provide. Based on measures and comparisons over time or between practices, suggestions can be made to improve the quality of chronic care management.

AUDIT AND FEEDBACK IN AN ACCREDITATION PROGRAM

Audit and feedback can be used to assess and improve the quality of chronic care management and to set up tailored improvement programs. The key element of audit and feedback is the provision of 'any summary of clinical performance of healthcare over a specified period of time, given in a written, electronic or verbal format'[16]. This summary is often based on performance measurement with use of indicators. It can also contain a comparison with other performance scores or a benchmark. Ivers et al.[17] performed a systematic review on the effects of audit and feedback. They found that there are small but potentially important improvements in care, including chronic care, albeit with large variation in effects across the trials. These effects are possibly larger when baseline performance is low, when feedback is provided both verbally and as text and includes explicit targets and an action plan.

Accreditation can be seen as a special type of audit and feedback and is widely used as a strategy to improve quality of chronic care management[18]. A closely related term that is often used interchangeably is certification. Accreditation affects the institution or practice and is offered more or less voluntary, while certification focuses on a specific norm that should be reached by individuals or particular services[19,20]. Greenfield et al.[21] found in their systematic review that accreditation can promote change, for example through the opportunity to reflect on organizational performance, and that it has an effect on professional development.

In the Netherlands, a practice accreditation program has been set up to evaluate general practices regarding chronic care management, practice organization and patient experiences with care; see Box 1 for further details. A crucial feature of this program is that it is primarily aimed at stimulating quality improvement in participating practices. This thesis focuses solely on the components in the Accreditation program that regard chronic care management.

Box 1. The Dutch Accreditation program

In the Netherlands, general practices can take part in a Dutch Accreditation program since 2005[22]. This program strongly focuses on the educational value of audit and feedback and participation is voluntary. However, the program does require that specific norms are met, for example on hygiene, first aid equipment and accessibility. The program evaluates practice organization, patient opinions and chronic care management. The structured program with a three-year cycle starts with a preparatory phase which consists of the collection of data on practice management and patient care. The measurement instruments used are previously validated questionnaires, including the 'VIP', a visitation instrument for practice organization[23] and the 'Europep' that measures patient experiences[24]. The questionnaires are filled in by family physicians, nurses and patients. Furthermore, a trained observer pre-audits the practice. Clinical performance is measured with the use of patient information that is extracted from electronic medical records; the family physician or nurse extracts the information either automatically or manually with an extraction form. When all data are collected and submitted through an online questionnaire system, the practice receives a feedback report that includes information on its own performance and the performance of other general practices as benchmarks. This information helps to identify which areas could be improved upon. The physicians then write improvement plans with a plan-do-study-act cycle. The first audit is carried out after the approval of these plans to confirm adequate participation and to grant accreditation. After this audit a three-year accreditation cycle starts. At the end of each year the practice staff evaluates whether the objectives of improvement programs are met and writes new improvement programs for the following year. The prolongation of the accreditation depends on this process. Accreditation is not based on the actual quality of care itself but rather on the quality of the improvement initiatives. After three years, a new cycle starts with the data collection phase.

PERFORMANCE MEASURES

In most accreditation programs, performance measures are used to evaluate individual practices[25]. These performance measures need to be tested empirically on their metric properties, including validity, reliability and feasibility, before broad implementation[26,27]. Validity and feasibility of performance measures from other programs have been described in international literature[28,29]. Regarding the measurement instruments used to evaluate chronic care management in the Dutch Accreditation program, content validity and clinical relevance have been assessed during the development of the performance measures used in the Dutch Accreditation program[22]. However, other metric properties have not been examined. In *chapter 2*, we focus on several aspects of validity and reliability: correlation between different sets of indicators, internal consistency within each set and precision of the performance scores and benchmarks. Types of reliability that are frequently examined in the literature include internal consistency and test-retest reliability[30,31]. However, an important aspect of reliability that has not been described extensively is precision. Precision relates to whether the estimated performance score is an adequate reflection of the population value[32]. This depends on several aspects such as the variation of scores and the number of data points used to calculate performance scores. In *chapter 3* we investigate whether performance measures used in the Dutch Accreditation program are based on samples of patients which are sufficiently large to ensure precision. Which level of precision is acceptable for routine use partly depends on the goal of the measurement: discriminative or evaluative[33]. For

discriminative purposes, higher levels of precision are needed. One possible solution to achieve higher levels of precision is to create a composite performance measure based on a group of related measures. We therefore examine which consequences the creation of a composite measure has on the precision.

IMPACTS OF ACCREDITATION

Alkhenizan et al.[19] performed a systematic review on the effects of accreditation programs. They found that accreditation had positive effects on processes as well as on clinical outcomes; they concluded that accreditation can be an important tool to improve quality of care. However, this review focused on accreditation in hospital care; few rigorous evaluations of effectiveness of primary care accreditation are available[34,35]. O'Beirne et al.[36] performed a review of peer-reviewed and grey literature regarding accreditation in primary care. They found indications that accreditation may improve both organization and outcomes of care, but conclude that more research is needed. In *chapter 4*, we examine the effects of the Dutch Accreditation program on chronic care management in general practice.

Insight in participant experiences with practice accreditation can help to identify new approaches to improving practice accreditation[37]. Therefore, we planned a qualitative study in participating practices to evaluate the Dutch Accreditation program, which is described in *chapter 5*.

A vital component of the Accreditation program consists of the provision of feedback to the practice. However, in order to derive meaningful interpretations from the performance scores, it is important to be aware of factors that may influence these scores. Previous research showed that health outcomes of an individual patient are affected by a range of non-modifiable factors that include age, gender and the duration of the chronic condition[38-40]. Multi-morbidity, defined as any co-occurrence of medical conditions within a person[41], can also affect chronic care processes and outcomes[42]. Depending on the goal of measurement and strength of the associations found, one can choose to adjust or stratify for these influencing factors[43]. In *chapter 6*, we investigate whether several patient characteristics (age, gender, body mass index (BMI), whether the patient lives in a deprived area) and multi-morbidities (cardiovascular disease, COPD, depression and anxiety) are associated with diabetes care processes and outcomes.

When the current state of care is evaluated, the next step is to elaborate targeted improvement plans and implement changes in the practice. Many interventions to improve the quality of care are aimed at practice management aspects[44]. The Chronic Care Model[9] described earlier offers a useful framework for the domains on which these interventions can be focused. Tricco et al.[45] performed a systematic review and provided a model that incorporates interventions focused on 12 different target areas. They found that interventions

aimed at more than one target area are more effective. However, the studies included in this review are controlled trials. It is uncertain whether similar effects can be found in routine primary care. In *chapter 7*, we examine which practice characteristics linked with intervention strategies are associated with favorable chronic care management. Table 1 presents the research questions of studies in this thesis.

Table 1. Overview of research questions of studies in this thesis

Chapter	Research questions
Chapter 2	To which extent does the domain clinical care of the Visitation Instrument Accreditation (VIA) used in the Dutch Accreditation program meet each of four reliability and validity criteria? 1. Correlation between the sets of performance indicators 2. Internal consistency within each set of performance indicators 3. Precision of indicator scores 4. Precision of benchmarks
Chapter 3	What is the relationship between sample size and precision? Can we increase precision with a composite performance score? How many indicators are needed minimally to achieve a certain level of precision?
Chapter 4	What is the effect of the Dutch Accreditation program on performance scores regarding diabetes, COPD and cardiovascular disease?
Chapter 5	What are the opinions of family physicians regarding the Dutch Accreditation program?
Chapter 6	Which patient characteristics are associated with performance scores?
Chapter 7	Which organizational factors are determinants of intermediate patient outcomes in routine diabetes care?

REFERENCES

1. Epping-Jordan J, Bengoa R, Kawar R, Sabate E. The challenge of chronic conditions: WHO responds – The sooner governments act, the better. BMJ 2001;323(7319):947-8.
2. Yach D, Hawkes C, Gould CL, Hofman KJ. The global burden of chronic diseases: overcoming impediments to prevention and control. JAMA 2004;291(21):2616-22.
3. Bodenheimer T, Chen E, Bennett HD. Confronting the growing burden of chronic disease: can the U.S. health care workforce do the job? Health Aff 2009;28(1):64-74.
4. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007;22:391-5.
5. Cowie CC, Rust KF, Byrd-Holt DD, Eberhardt MS, Flegal KM, et al. Prevalence of diabetes and impaired fasting glucose in adults in the U.S. population: National Health and Nutrition Examination Survey 1999-2002. Diabetes care 2006;29(6):1263-8.
6. Fang J, Shaw KM, Keenan NL. Prevalence of coronary heart disease – United States, 2006-2010. JAMA 2011;306(19):2084-6.
7. Murray-Thomas T, Cowie MR. Epidemiology and clinical aspects of congestive heart failure. J Renin Angiotensin Aldosterone Syst 2003;4(3):131-6.
8. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. Lancet 2007;370(9589):741-50.
9. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-44.
10. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002;288(14):1775-9.
11. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the Chronic Care Model in the new millennium. Health Aff 2009;28(1):75-85.
12. van Lieshout J, Goldfracht M, Campbell S, Ludt S, Wensing M. Primary care characteristics and population-orientated health care across Europe: an observational study. Br J Gen Pract 2011;61(582):e22-e30.
13. Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, et al. Substitution of doctors by nurses in primary care. Cochrane Database Syst Rev 2005;2:CD001271.
14. Martinez-Gonzalez NA, Djalali S, Tandjung R, Huber-Geismann F, Markun S, et al. Substitution of physicians by nurses in primary care: a systematic review and meta-analysis. BMC Health Serv Res 2014;14:214.
15. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005;330(7494):765.
16. Jamtvedt G, Young JM, Kristoffersen DT, Thomson O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2003;3:CD000259.
17. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012;6: CD000259.
18. Shaw C. The external assessment of health services. World Hosp Health Serv 2004;40(1): 24-7, 50, 51.
19. Alkhenizan A, Shaw C. Impact of accreditation on the quality of healthcare services: a systematic review of the literature. Ann Saudi Med 2011;31(4):407-16.
20. Shaw C, Groene O, Mora N, Sunol R. Accreditation and ISO certification: do they explain differences in quality management in European hospitals? Int J Qual Health Care 2010;22(6):445-51.
21. Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review. Int J Qual Health Care 2008;20(3):172-83.
22. Braspenning J, Dijkstra R, Tacken M, Bouma M, Witmer H. Visitatie Instrument Accreditering (VIA®). Nijmegen/Utrecht: Kwaliteit van Zorg (WOK); UMC St Radboud; Nederlands Huisartsen Genootschap (NHG); NHG Praktijk Accreditering BV (NPA); 2007.
23. van den Hombergh P, Grol R, van den Hoogen HJ, van den Bosch WJ. Assessment of management in general practice: validation of a practice visit method. Br J Gen Pract 1998;

48(436):1743-50.
24. Grol R, Wensing M, Mainz J, Jung HP, Ferreira P, et al. Patients in Europe evaluate general practice care: an international comparison. Br J Gen Pract 2000;50(460):882-7.
25. Braspenning J, Hermens R, Wollersheim H, Grol R. Meten van (veranderingen in) de zorg: de rol van indicatoren. In: Grol R, Wensing M (redactie). Implementatie: effectieve verbetering van de patiëntenzorg. Maarssen: Elsevier gezondheidszorg; 2006 (3e druk): 153-79.
26. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. BMJ 2003;326(7393):816-9.
27. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2003 (3rd ed.).
28. Campbell SM, Hann M, Hacker J, Durie A, Thapar A, et al. Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes. Qual Saf Health Care 2002;11(2):125-30.
29. Campbell SM, Roland MO, Quayle JA, Buetow SA, Shekelle PG. Quality indicators for general practice: which ones can general practitioners and health authority managers agree are important and how useful are they? J Pub Health Med 1998;20(4):414-21.
30. Steinhaeuser J, Miksch A, Ose D, Glassen K, Natanzon I, et al. Questionnaire of chronic illness care in primary care – psychometric properties and test-retest reliability. BMC Health Serv Res 2011;11:295.
31. Parkerton PH, Smith DG, Belin TR, Feldbau GA. Physician performance assessment: nonequivalence of primary care measures. Med Care 2003;41(9):1034-47.
32. Bernstein S, Bernstein R. Schaum's outline of elements of statistics 2: inferential statistics. New York, NY: McGraw-Hill; 1999.
33. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? J Clin Epidemiol 1992;45 (12):1341-5.
34. Greenfield D, Braithwaite J. Developing the evidence base for accreditation of healthcare organisations: a call for transparency and innovation. Qual Saf Health Care 2009; 18(3):162-3.
35. Hinchcliff R, Greenfield D, Moldovan M, Westbrook JI, Pawsey M, et al. Narrative synthesis of health service accreditation literature. BMJ Qual Saf 2012;21(12):979-91.
36. O'Beirne M, Zwicker K, Sterling PD, Lait J, Lee Robertson H, et al. The status of accreditation in primary care. Qual Prim Care 2013;21(1):23-31.
37. Greenfield D, Pawsey M, Braithwaite J. What motivates professionals to engage in the accreditation of healthcare organizations? Int J Qual Health Care 2011;23(1):8-14.
38. Schaars CF, Denig P, Kasje WN, Stewart RE, Wolffenbuttel BH, et al. Physician, organizational, and patient factors associated with suboptimal blood pressure management in type 2 diabetic patients in primary care. Diabetes care 2004;27(1):123-8.
39. Campbell JA, Walker RJ, Smalls BL, Egede LE. Glucose control in diabetes: the impact of racial differences on monitoring and outcomes. Endocrine 2012;42(3):471-82.
40. Strom Williams JL, Lynch CP, Winchester R, Thomas L, Keith B, et al. Gender differences in composite control of cardiovascular risk factors among patients with type 2 diabetes. Diabetes Technol Ther 2014;16(7):421-7.
41. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: what's in a name? A review of literature. Eur J Gen Pract 1996;2(2):65-70.
42. Turner BJ, Hollenbeak CS, Weiner M, Ten Have T, Tang SS. Effect of unrelated comorbid conditions on hypertension management. Ann Intern Med 2008;148(8):578-86.
43. Chin MH. Risk-adjusted quality of care rating for diabetes: ready for prime time? Diabetes care 2000;23(7):884-6.
44. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. Prev Chronic Dis 2013;10:E26.
45. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet 2012;379(9833):2252-61.

2

Evaluation of clinical indicators: four reliability and validity issues

Arna van Doorn
Kirsten Kirschner
Margriet Bouma
Jako Burgers
Jozé Braspenning
Richard Grol

Published as: 'Evaluatie van het onderdeel medisch handelen van de accreditering. Vier klinimetrische criteria' in Huisarts en Wetenschap 2010; 53(3): 141-6.

ABSTRACT

Background To describe four reliability and validity issues regarding the clinical indicators from the Visitation Instrument Accreditation (VIA). Based on this information, practices needed to develop and implement improvement plans in order to get accredited.

Methods An observational study based on data from the medical records of 82 practices.

Results The performance indicators that covered chronic care management (diabetes, COPD, asthma and cardiovascular risk management), prevention activities (influenza vaccination, cervical cancer screening) and prescription of antibiotics were correlated weakly, suggesting that the instrument provided a rather broad scope of the practice when it comes to chronic care management and prevention. Furthermore, the different subjects were each measured by indicators that had sufficient coherence, which suggested that they measured a clear underlying concept. To achieve a high indicator score precision (i.e. a confidence limit of plus/minus 10 percentage points), data from at least 96 patients were necessary to calculate a score. VIA allows to take a sample of 40 patients; in that case the confidence limit increases to 15 points. To establish a precise benchmark we needed 233 practices to achieve a confidence limit of ≤ 5 percentage points.

Conclusions The clinical indicators of the VIA are reliable and valid and can be used by a general practice to gain insight into its performance compared to a benchmark. For practice policy on quality improvement a confidence limit of 10-15 percentage points around the indicator score on practice level seems to be acceptable. We would be more comfortable with a smaller confidence limit in case of accountability or a 'pay-for-performance' program. A sample of 40 patients would not do in such a case. It therefore remains a continuous search to find a balance between feasibility and justice.

BACKGROUND

Clinical performance indicators are needed to measure the quality of care in order to create more transparency in healthcare[1]. To achieve this, general practices can use the Visitation Instrument Accreditation (VIA), which maps various aspects of general practice care with several indicator sets[2]. Since 2005, the VIA has been used in the Dutch Accreditation program of the Dutch College of General Practitioners, see Box 1. The instrument consists of the following domains: clinical care, practice management and patient experiences. Indicators are used to gain more insight into the quality of care, and enable each family physician to compare the performance of their practice to a benchmark that is based on performances of all participating practices. This feedback method offers family physicians guidance to improve the quality of their care. This type of information is increasingly used, for instance for internal policy making, multidisciplinary care groups and public information. Because of this, questions regarding the validity and reliability of this information are of notable importance.
The validity and reliability of the domains practice management and patient experiences have been described before[3,4]. However, little is known about the clinical care domain. During the development of these performance measures, content validity has been assessed for the individual indicators: a panel of experts agreed that the indicators are a good reflection of clinical care and can be used to evaluate this domain[2,5]. Now that data have been collected with the instrument, it is also possible to examine other aspects of reliability and validity[6,7].

Box 1. The Dutch Accreditation program (NHG-Praktijkaccreditering®)

Since 2005, family physicians (FPs) can participate in the Dutch Accreditation program of the Dutch College of General Practitioners (NHG). FPs collect data for their accreditation using the Visitation Instrument Accreditation (VIA). The VIA consists of performance indicators with regard to clinical care, practice management and patient experiences, and measures the quality of (part of) the activities carried out in the general practice. Data are extracted from medical records and questionnaires. In addition, observations are made by a consultant of NPA, the organization that carries out the independent review on behalf of the Dutch Accreditation program. On the basis of the incorporated data the practice is able to compare itself to other practices, and can consequently draw up plans for improvement. Accreditation of the practice is based on the development and implementation of these improvement plans, provided that certain conditions (minimum requirements) have been met. This is checked by the NPA accreditor during an audit. Data are collected using the VIA in the first year. In the following two years further audits of the practice take place, in which the accreditor checks whether the physicians have achieved to implement the plans from that year, whether the improvement plans for the upcoming year meet the requirements and whether the minimum requirements for that particular year have been met. The cycle repeats itself after three years.

This study examines clinical indicators on the basis of four reliability and validity issues. First, the various sets of indicators of the instrument should describe a broad and diverse range of clinical activities, which take place in a general practice. This means that the sets of indicators should not be too closely connected to each other (low correlation between the sets of

indicators). Second, it is important that there is a certain amount of coherence between the indicators within one set, though the information should not overlap too much (internal consistency within a set). Third, the performance score that is calculated from the indicator should give an accurate picture of the practice. This means that data should be collected from a sufficient number of patients (precise indicator score). Fourth, in order to provide a stable and correct reflection of the scores in a population, a benchmark should be based on a sufficient number of practices. Moreover, the practices on the basis of which the benchmark is calculated, should be representative of the practice that uses the benchmark (adequate and precise benchmark). We studied the extent to which the domain clinical care of the VIA meets each of these criteria [2].

METHODS

Study population and measures

We carried out an observational study based on data from the medical records of 82 practices that voluntarily took part in the Dutch Accreditation program during 2005-2006. The practices collected data on three widespread chronic conditions (diabetes, chronic obstructive pulmonary disease (COPD) and asthma), cardiovascular risk management (CVRM), a number of specific prevention activities (influenza vaccination and cervical cancer screening) and prescription of antibiotics. The instrument consisted of structure indicators, process indicators and outcome indicators[8]. We mostly concentrated on the process indicators for the examination of the criteria, because these are the most reliable in providing information on the quality of practice management and care. Table 1 shows an overview of all indicators that were included in our study. We used 9 process indicators for diabetes, 8 for CVRM, 5 for COPD, 4 for asthma, and 1 indicator each for influenza vaccination, cervical cancer screening and prescription of antibiotics. We also included 5 outcome indicators (3 diabetes and 2 CVRM indicators) for a number of the calculations, in order to gain insight in the effect of this type of information on the reliability and validity.

Data collection

The required data were collected from the medical records by a staff member of the general practice. However, it was often not easy to extract data from the electronic medical records; for one thing because the software did not offer easy data extraction, but also because data had not been registered in a uniform way. Therefore, the practices were offered the choice to collect data from a sample of 40 patients. If the total number of patients with the condition concerned did not exceed 40, the practices were requested to collect data from all patients.

Table 1. Overview of all indicators

Subject	No.	Description
Diabetes	1	% of patients with completed information on risk factors
Process indicators	2	% of patients with a record of at least 3 glucose measurements in the last 12 months
	3	% of patients with a record of HbA1c measurement in the last 12 months
	4	% of patients with a record of blood pressure measurement in the last 12 months
	5	% of patients with a record of total cholesterol measurement in the last 12 months
	6	% of patients with a record of creatinine measurement in the last 12 months
	7	% of patients with a record of retinal examination in the last 24 months
	8	% of patients with a record of feet examination in the last 12 months
	9	% of patients with a record of a lipid lowering medication prescription
Outcome indicators	1	% of patients with an Hba1c value of 8.5% or more
	2	% of patients with a blood pressure value within target level (150/85 mm Hg or less)
	3	% of patients with a total cholesterol value within target level (5.0 mmol/l or less)
COPD	1	% of patients with a record of a spirometry assessment ever
	2	% of patients with a record of a spirometry assessment in the last 12 months
	3	% of patients with a record of a consultation in the last 12 months
	4	% of patients with a record of smoking status
	5	% of patients who smoke with a record of smoking cessation advice
Asthma	1	% of patients with a record of a spirometry or peakflow measurement ever
	2	% of patients with a record of a consultation in the last 12 months
	3	% of patients with a record of smoking status
	4	% of patients who smoke with a record of smoking cessation advice
CVRM	1	% of patients with a record of blood pressure measurement in the last 12 months
Process indicators	2	% of patients with a record of cholesterol measurement in the last 12 months
	3	% of patients with a record of a lipid lowering medication prescription
	4	% of patients with a record of smoking status
	5	% of patients who smoke with a record of smoking cessation advice
	6	% of patients with completed information on risk factors
	7	% of patients with heart disease in their history with a record of a prescription of anticoagulants
	8	% of patients with a record of a glucose measurement in the last 12 months
Outcome indicators	1	% of patients with a blood pressure value within target level (160/90 mm Hg or less)
	2	% of patients with a lipid lowering medication prescription with a total cholesterol value within target level (5.0 mmol/l or less)
Influenza vaccination	1	% of vaccinated high risk patients in the practice or % of vaccinated patients of 65 years and older*
Cervical cancer screening	1	% of women from the target cohort with a record of a cervical smear
Antibiotics	1	% of narrow-spectrum antibiotic prescriptions in relation to all prescriptions of antibiotic drugs

** We decided to use the highest value of these two indicators to serve as indicator score for this item, because practices were often not able to fill in both indicators.*

The Visitation Instrument Accreditation (VIA) requested details to be collected only for that part of the patient group that is treated by the family physician (FP). However, reports were often made on all patients with the condition concerned. With regards to the prescription of antibiotics data, practices were advised to contact the preferential pharmacist. The data collection could refer to just one FP, or to more FPs if they shared the patient population. Of the 82 practices, we included a total of 97 patient populations; for each population we had data on all conditions included in the study.

Analysis

In order to find out whether the practices in our study were representative of all Dutch practices, we compared our study population to all Dutch general practices concerning practice type, degree of urbanization and whether practices had a dispensary. Moreover, we looked at the number of patients per full-time equivalent (FTE) FP.

To investigate whether the measurements of the various conditions of clinical care from the VIA differed enough from each other, we determined the level of correlation between the various subjects (criterion 1) using Pearson's correlation coefficient. To assess the internal consistency (criterion 2) we determined the correlation between the various indicators within one subject by using Cronbach's alpha[9]. A Cronbach's alpha between 0.7 and 0.8 is advisable, although an alpha higher than 0.6 is also considered to be acceptable for these types of indicators[10]. For each indicator and patient population we calculated the number of patients needed in order to achieve a precise score (criterion 3). This calculation was based on the indicator score we found. We kept to a 95% significance level and tested for two separate confidence limits, of 5 percentage points and 10 percentage points. For each indicator we then calculated the maximum, the 75th percentile score, the mean and the standard deviation, in order to get a clear picture of the scores and the levels of dispersion. We also used a power calculation to estimate the number of patient populations needed to get a precise benchmark (criterion 4)[11]. For this, we also used a 95% significance level and tested for the two confidence limits of 5 and 10 percentage points.

RESULTS

Study population

Most characteristics of the general practices in our study reasonably correspond to the characteristics of all Dutch general practices (Table 2). Yet there seem to be relatively fewer single handed practices in our study. Also, the number of patients per FTE FP is higher in our study population than the national number of residents per FTE FP.

Table 2. Practice characteristics of the study population in comparison with all Dutch practices (2008)

Practice characteristics	Study population Dutch Accreditation program, n = 82 practices		All Dutch general practices* n = 4235 practices
	n	*%^*	*%^*
Practice type			
Single handed practice	18	22.0	42.3
Practice with two FPs	18	22.0	31.5
Group practice or health center	24	29.3	26.1
HOED◊	15	18.3	-
Other type of practice	7	8.5	-
Practice location‡			
Urban	33	40.2	46.7
Partly urban	38	46.3	41.1
Rural	11	13.4	12.2
Dispensary			
Yes	3	3.7	7.3
No	79	96.3	92.7
No. of patients per FTE FP	2444		
No. of inhabitants per FTE FP			2322

** Data provided by NIVEL, 1-1-2008[12]. ◊ HOED: construction where multiple family physicians (FPs) have their separate practices under one roof. ‡ A region was defined as urban when the number of addresses per km² exceeded 1,500; partly urban regions had 500 - 1.500 addresses per km²; rural regions had less than 500 addresses per km². ^ Because the percentages were rounded off, they do not always add up to exactly 100%.*

Criterion 1: low correlation between the indicator sets

We did not find strong correlations between the various subjects of clinical care (see Table 3). The subject prescription of antibiotics did not show coherence with any of the other subjects. Although cervical cancer screening turned out to be slightly associated with cardiovascular risk management, this correlation was rather weak. Of the other 5 subjects, asthma showed the strongest association with the other subjects (diabetes, COPD, cardiovascular risk management and influenza vaccination). Furthermore, we noticed a slight correlation between diabetes and both cardiovascular risk management and influenza vaccination.

Criterion 2: internal consistency of the indicator sets

Table 4 shows the average scores on the indicators for each subject, the standard deviation and the internal consistency of the indicators for each subject (Cronbach's alpha). The 9 process indicators which relate to diabetes proved to have the highest internal consistency. These indicators had a Cronbach's alpha of 0.73. This score can vary between 0 and 1, with a score of 0 reflecting no inter-correlation at all and 1 reflecting a perfect inter-correlation. When a score is close to 0.9 or higher, the internal consistency is so strong that there is an overlap in the information provided by the indicators. An option then would be to leave out

indicators. Scores closer to 0 suggest that the indicators deal with different subjects, which means that the indicators ought not to be taken together. The 8 process indicators relating to cardiovascular risk management had a reasonably strong internal consistency, just as the 5 indicators relating to COPD (respective alphas of 0.64 and 0.67). The internal consistency of the 4 asthma indicators proved to be slightly less strong (alpha = 0.56). Describing the internal consistency using Cronbach's alpha is only possible for subjects with more than one indicator.

Table 3. Correlation between the subjects of clinical care (Pearson's correlation coefficient of the means)

Subject	Diabetes	COPD	Asthma	CVRM	Influenza	Cancer Screening	Antibiotics
Diabetes	X	0.15	0.31*	0.43*	0.29*	0.19	-0.10
COPD		X	0.40*	0.09	0.20	-0.13	0.12
Asthma			X	0.48*	0.30*	0.16	0.06
CVRM				X	0.08	0.27*	0.19
Influenza vaccination					X	0.17	0.17
Cancer screening						X	-0.07
Antibiotics							X

** Significant correlation, p <0.01.*

Table 4. Internal consistency within each subject (Cronbach's alpha, n = 97)

Subject	No. of indicators	Cronbach's alpha	Mean in %	SD in %
Diabetes	9	0.73	72.4	11.2
COPD	5	0.67	63.1	15.4
Asthma	4	0.56	49.3	17.1
CVRM	8	0.64	49.6	11.4
Influenza vaccination	1	-	85.6	8.4
Cancer screening	1	-	67.6	16.5
Antibiotics	1	-	14.3	11.5

Criterion 3: precise performance scores per patient population

Using the resulting performance scores, we calculated the number of patients needed to achieve a certain level of precision for each patient population and indicator separately. Table 5 lists the minimum number of patients needed per subject. When subjects consisted of more than one indicator we calculated the numbers for each indicator separately; in that case Table 5 lists the lowest and the highest number of patients needed for each subject. Looking at the achieved performance scores, we needed at least 363 and at most 384 patients to calculate a precise score for the 9 diabetes process indicators, with a confidence limit of plus or minus 5 percentage points. For many of the subjects, at least 384 patients were needed. If we would accept a larger confidence limit of 10 points, we would need to assess data from 96

patients to achieve a precise score. The average population size of the participating practices was 3,917 patients. This means that we can only calculate precise indicator scores (with a confidence limit of ≤ 10 points) for conditions with a minimal prevalence of 2.4%. This is the case for diabetes, asthma and cardiovascular risk management (RIVM site[13]). The prevalence of COPD is just below 2.4%. It is probable that not all patients with this condition are registered in general practices, which means that through an improved registration we will also be able to include enough patients for COPD to achieve precise indicator scores.

Table 5. Number of patients/patient populations needed to achieve 5 or 10 point confidence limits around indicators and benchmarks (highest and lowest for each subject)

Subject (no. of indicators)	No. of patients needed for each indicator		No. of populations needed for each benchmark	
	CL 5 points*	*CL* 10 points*	*CL* 5 points*	*CL* 10 points*
Diabetes (p◊, 9)	363-384	91-96	8-233	2-58
Diabetes (o◊, 3)	272-384	68-96	4-35	1-9
COPD (5)	384-384	96-96	46-146	11-37
Asthma (4)	384-384	96-96	64-175	16-44
CVRM (p◊, 8)	384-384	96-96	28-161	7-40
CVRM (o◊, 2)	384-384	96-96	39-46	10-11
Influenza vaccination (1)	372	93	11	3
Cervical cancer screening (1)	384	96	42	10
Antibiotics (1)	383	96	20	5

** CL = confidence limit. ◊ p = process indicators, o = outcome indicators.*

Criterion 4: precise benchmark

The number of patient populations needed to calculate a precise benchmark depends on the dispersion of the various scores on the indicator concerned. The standard deviation for each indicator varied between 5.1 for the diabetes outcome indicator 'patients with an Hba1c value of 8.5% or more' and 39.0 for the diabetes process indicator 'patients with completed information on risk factors'. Due to the fact that there is a strong dispersion between the various indicators, there are also large differences in the number of populations needed for each indicator. For instance, the minimum number of patient populations needed for the diabetes indicators varies between 8 and 233 with a confidence limit of 5 points (Table 5). To calculate a precise benchmark for all indicators, 233 patient populations are needed for a confidence limit of 5 points and 58 populations for a confidence limit of 10 points. The benchmarks that are currently used as reference material are based on 259 populations, which is more than sufficient to maintain a maximum confidence limit of 5 percentage points.

DISCUSSION

The results show that the domain of clinical care, as part of the Visitation Instrument Accreditation, consists of a relatively broad and diverse scope of care in general practice. The associations between the conditions that have been included are relatively weak. The correlations that we found can be explained by the overlap between the various groups. For example, a patient who has been included in the indicators concerning cardiovascular risk management could also suffer from diabetes. Furthermore, such a patient would also qualify for an influenza vaccination. In the future, more subjects can be added to the instrument in order to make the evaluation more complete and cover more parts of the field of primary care. The subjects that are discussed in the present version are assessed by means of a set of indicators which show a reasonable internal consistency. The exception was the Cronbach's alpha score relating to asthma, which showed a rather low correlation. A possible explanation for a lower correlation between asthma indicators is the fact that the guidelines for check-up consults are less strict for asthma, especially when the patient is on little or no medication. A reasonable to good association between indicators serves as an indication that we are measuring one underlying concept. With respect to the amount of patients needed to establish precise scores, often the same maximum number of patients emerged for the various indicators. The reason for this was that with almost all indicators, at least one patient population had a score of 50%; with this score the largest number of patients is statistically needed to achieve a precise score. The number of patients needed decreases the closer a score gets to 0% or 100%. The reliability of the scores is also dependent on the reliability of the data that are used for their calculation. The scores will never be reliable when the data have not been collected in a reliable way, for instance because of problems with the data collection. Uniform reporting is therefore of the utmost importance. Moreover, the data that will be included or excluded should be determined in advance.

The question of how many patients are needed to calculate a precise score and which confidence limit would be acceptable, depends to a large extent on the purpose of the measurement. The score does not have to be accurate to one percent in order to give a good indication of possible points for improvement if it is used for internal quality improvements only. Drawing a line with respect to which accuracy level is acceptable is rather arbitrary and partly determined by feasibility factors. A general practice only has a limited number of patients with a certain condition. There is also a limit to the time investment by the family physician. The sample survey of 40 patients which is permitted at the moment in the VIA is reliable as long as a confidence limit of 15 percentage points is accepted. For want of larger numbers of data, these scores could indeed be used as a reasonable indication for internal quality improvement. Stricter demands are put upon the instrument when it should be able

to differentiate between practice scores or when it should report scores more accurately. If the instrument is used to distinguish between practices, for instance as part of a pay-for-performance-system, larger numbers of patients will have to be included. A confidence limit of 5 percentage points or less is advisable for this purpose, which means a distinction can only be made on the level of the care group or between larger practices.
In our study we did not calculate performance scores on a practice level, but on the level of the patient population. In most cases a practice contained only one population. When practices supply data for several populations, these data can be taken together in order to give a better picture of the practice as a whole. When interpreting data it should be taken into consideration that the division between 'reliable' and 'not reliable' is not as clear as it might seem. Even if, on the basis of the number of patients included, the conclusion is that an indicator score is not precise enough, such a score could still give a reasonably good picture of the practice. However, in this case we are not 95% sure that the score is indeed representative, meaning that care must be taken when conclusions are drawn on the basis of the scores found. The number of patients or patient populations needed will decrease by raising the confidence limit to 15 points. Depending on the purpose, the confidence limit chosen and the number of patients included can be varied upon in order to arrive at a score which is as precise as possible.

Conclusions

Our study shows that the domain of clinical care, part of the VIA, is reliable and valid with regard to the four criteria studied. However, it is important in this respect to take an acceptable confidence interval into account when interpreting data. Uniform reporting is of the utmost importance to improve the reliability of the data collection. The instrument is valuable for FPs to gain more insight into the medical care they provide, compared with national reference figures. On the basis of this instrument, FPs can formulate plans to improve the quality of clinical care. The instrument can also be useful when looking at the quality of clinical care on a national scale. However, the number of patients that should be included in the data collection would then need to be larger than for use of the instrument within the practice, due to the higher demands on precision.

REFERENCES

1. Grol R, Wensing M (redactie). Implementatie: effectieve verbetering van de patiëntenzorg. Maarssen: Elsevier gezondheidszorg; 2006 (3e druk).
2. Braspenning J, Dijkstra R, Tacken M, Bouma M, Witmer H. Visitatie Instrument Accreditering (VIA®). Nijmegen/Utrecht: Kwaliteit van Zorg (WOK); UMC St Radboud; Nederlands Huisartsen Genootschap (NHG); NHG Praktijk Accreditering BV (NPA); 2007.
3. Grol R, Wensing M, Mainz J, Ferreira P, Hearnshaw H, et al. Patients' priorities with respect to general practice care: an international comparison. European Task Force on Patient Evaluations of General Practice (EUROPEP). Fam Pract 1999;16(1):4-11.
4. van den Hombergh P, Grol R, van den Hoogen HJ, van den Bosch WJ. Assessment of management in general practice: validation of a practice visit method. Br J Gen Pract 1998; 48(436):1743-50.
5. Braspenning JCC, Pijnenborg L, In 't Veld CJ, Grol RPTM (redactie). Werken aan kwaliteit in de huisartsenpraktijk: indicatoren gebaseerd op de NHG-Standaarden. Houten: Bohn Stafleu van Loghum; 2005.
6. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. Qual Saf Health Care 2002; 11(4):358-64.
7. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2003 (3rd ed.).
8. Donabedian A. The definition of quality and approaches to its assessment (vol. 1): explorations in quality assessment and monitoring. Ann Arbor, MI: Health Administration Press; 1980.
9. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297-334.
10. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997;314(7080):572.
11. Moore DS, McCabe GP. Statistiek in de praktijk. Schoonhoven: Academic Service; 1999.
12. Hingstman L, Kenens RJ. Cijfers uit de registratie van huisartsen – peiling 2008. Utrecht: NIVEL; 2009.
13. website RIVM, nationaal kompas volksgezondheid. http://www.nationaalkompas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen.

3

Precision of individual and composite performance scores: the ideal number of indicators in an indicator set

Arna van Doorn - Klomberg
Jozé Braspenning
Remco Feskens
Margriet Bouma
Stephen Campbell
David Reeves

Medical Care 2013; 51(1): 115-21.

ABSTRACT

Background In many countries, quality indicators are used to assess the quality of care of general practices. Such assessments need to have an adequate precision, so that the results can be interpreted correctly. However, a small sample size per physician can lead to inadequate precision. A possible solution could be to create composite performance scores. The objective of this study is to evaluate the relationship between sample size and precision. We examine whether a composite performance score has an increased precision and how many indicators are needed minimally to achieve this level of precision.

Methods We performed a descriptive statistical study on data from the medical records of 455 Dutch practices. We included three different conditions: diabetes (12 indicators), chronic obstructive pulmonary disease (4 indicators), and cardiovascular disease and risk management (9 indicators).

Results For individual quality indicators, patient samples close to 100 are required to achieve even moderate precision (10 percentage points) on the performance scores. This number decreases substantially when a composite score is used. A composite derived from combining 5 to 7 indicators can provide much the same precision of measurement as one made up from a much larger number of indicators.

Conclusions The added value of a composite score depends on the a priori reasons for measuring quality. Our results indicate that especially for formative quality improvement a small number of carefully selected indicators can provide a sufficiently precise composite measure.

BACKGROUND

Quality indicators provide a tool to assess the quality of primary care; different instruments have been developed to measure medical care, practice organization, and patient opinions[1-9]. To ensure meaningful quality evaluations, transparency is critical and assessments using indicators should adhere to key attributes that address the properties of the instrument such as validity, feasibility, and reliability[10-14]. However, there has been limited investigation into statistical aspects of indicator scores such as precision, particularly where several indicators are combined into a 'composite' measure.

Precision is an expression of the closeness of an estimate to the population value[15] and is typically measured as the width of the confidence interval around the estimated value. A related property is the reliability. Reliability is usually defined in terms of measurement reproducibility and freedom from random error[16], but when operationalized in the form of the reliability coefficient, becomes a comparative measure. Thus, whereas precision focuses on the accuracy of a performance score for one provider irrespective of the scores for other providers, the reliability coefficient compares the precision of an individual provider's score to the variability in scores across providers, as a measure of how distinct the provider scores are from one-another. The narrower the variation in scores between providers, the more precise the individual provider scores need to be to achieve high reliability. Validity, as a third property of a measurement, is essentially concerned with the extent to which a score measures what it purports to measure[13]; however, the current paper is not concerned with questions of validity.

Precision and reliability provide different kinds of information about a quality score. For the current paper we will focus mainly on precision, which is the more relevant measure for determining if a provider genuinely falls below a benchmark. Where providers receive a financial incentive based on their performance scores, poor precision could result in considerable overpayment or underpayment. Reliability is more informative where the aim is to ensure that the precision of measurement is sufficient to distinguish between different providers.

The precision of a score is dependent on the number of cases included in the sample. This poses a problem for routine clinical usage because the larger the patient sample, the more time consuming the data collection becomes. Furthermore, the achievable sample size, and hence the precision of measurement, depends on the prevalence of the condition and the size of the unit of assessment. The larger the unit, such as a region, the easier it is to access sufficient data to reach a desired level of precision. However, data at the individual physician level are more informative for Continuing Professional Development, and at the practice level for external quality assessment and formative practice level quality improvement[17]. Therefore,

it is important to investigate methods that help create more precise scores at finer levels of measurement in order for the scores to have clear interpretation (interpretability) and to distinguish one score from another score or a benchmark (discriminative ability).

A strategy that might increase precision at finer levels of measurement is to combine data on a number of individual quality indicators into a single 'composite' score. Reeves et al.[18] described different methods for creating composite scores and the different benefits and drawbacks of each, depending on the objective. Scholle et al.[19] looked at the reliability of individual versus composite scores and concluded that composite scores can increase reliability. Kaplan et al.[20] looked at the internal consistency of a composite score and found that adequate internal reliability could be achieved when at least 25 patients per physician were included.

This study investigates the relationship between the precision of indicator scores and the number of cases used for assessment and also how the latter is affected by creating composite scores. We will focus on diabetes, chronic obstructive pulmonary disease (COPD), and cardiovascular diseases, which are conditions that are often included in instruments that measure quality of primary care, such as the Healthcare Effectiveness Data and Information Set (HEDIS) in the United States[5] and the Quality and Outcomes Framework (QOF) in the United Kingdom[1]. These conditions have a high incidence and prevalence and account for high levels of expenditure and burden on individuals. As the number of quality indicators per condition varies across instruments, we will also address the influence of the number of indicators on the precision of the composite score. If we can achieve similar precision for a composite score based on a smaller subset of indicators, this would lower the burden for users of performance scores.

We address the following research questions:

1. What is the precision of a single performance score given different sample sizes?
2. What is the precision of composite performance scores given different sample sizes?
3. What is the smallest subset of indicators that will provide a composite score of similar precision to the full indicator set?

METHODS

Study population

A descriptive statistical study was conducted in January 2009, using medical record data from 455 Dutch general practices out of the 4,235 practices in the Netherlands[21]. The practices had taken part in the Dutch Accreditation program[8], on a voluntary basis, and had finished data collection between November 2005 and May 2008. The practices were a representative sample of all Dutch practices regarding practice size, location, and urbanization although

solo practices were slightly underrepresented in our sample. Practice staff collected data from electronic medical records using a standardized data abstraction form, taking a sample of 40 patients per practice. We removed patients where the medical record was entirely empty or the family physician was not the main care provider (see Appendix, excluded data). If, after these exclusions, the number of remaining patients for a condition at a practice was <20, we excluded the entire practice from the dataset for the condition. This resulted in a subset of 350 practices for diabetes, 286 practices for chronic obstructive pulmonary disease (COPD), and 342 practices for cardiovascular disease and risk management (CVD/RM).

Data collection

We extracted data from the patients' medical records relating to 12 clinical indicators on diabetes, 9 on CVD/RM, and 4 on COPD (Table 1). The CVD/RM indicators related to a heterogeneous group of people that had elevated risk for CVD: patients with hypercholesterolemia, hypertension, heart failure, angina pectoris, peripheral artery disease, or with myocardial infarction. Only process indicators were included in the study; such indicators are considered to be more sensitive to differences in quality of care compared with outcome indicators[22]. The indicators are based on Dutch guidelines and were developed using a consensus technique to ensure high content validity[8,23]. We required that there be a clear record of the indicator having been met in the patient's medical record, otherwise we treated it as not met.

Measures

For each indicator, a performance score was computed as the number of patients for whom the indicator was met, divided by the number of patients eligible for the indicator, expressed as a percentage; summary scores are listed in Table 1. A composite score was calculated for each of the three conditions as the mean of the individual indicator scores within the condition. This 'indicator average' method[18] forms the basis of many actual quality measurement systems[3,23]. Often these involve some form of weighting system to give differential importance to some indicators over others. The UK Quality and Outcomes Framework assigns weights to indicators based on expert judgement[1]. Another approach, termed 'opportunity weighting' gives each instance of an indicator equivalent impact on the composite score by weighting each indicator by its relative denominator (the denominator divided by the sum of denominators). The impact of weighting is not examined in detail in this paper, as this would add additional complexity to the analysis.

Table 1. Overview of indicators with median, 75th percentile and 90th percentile score across practices (diabetes n = 350, COPD n = 286, CVD/RM n = 342)

Condition	No.	Description	Median *(%)*	75th *(%)*	90th *(%)*
Diabetes	1	% of patients with a record of smoking status	70.5	89.7	97.6
	2	% of patients with a record of BMI*	75.7	94.4	100
	3	% of patients with a record of family history	29.9	68.9	94.9
	4	% of patients with a record of CVR◊ status	53.1	87.4	97.5
	5	% of patients with a record of at least 3 glucose measurements in the last 12 months	89.3	95.8	100
	6	% of patients with a record of HbA1c measurement in the last 12 months	95.8	100	100
	7	% of patients with a record of a blood pressure measurement in the last 12 months	97.1	100	100
	8	% of patients with a record of cholesterol measurement in the last 12 months	91.9	96.6	100
	9	% of patients with a record of creatinine measurement in the last 12 months	93.1	96.8	100
	10	% of patients with a record of retinal examination in the last 24 months	72.3	83.3	92.1
	11	% of patients with a record of feet examination in the last 12 months	62.3	81.7	91.7
	12	% of patients with a record of a lipid lowering medication prescription	61.3	72.2	79.5
COPD	1	% of patients with a record of a spirometry assessment ever	37.1	62.2	83.6
	2	% of patients with a record of spirometry assessment in the last 12 months	34.5	55.1	75.1
	3	% of patients with a record of smoking status	83.6	94.6	100
	4	% of patients with a record of a consultation in the last 12 months	73.9	84.6	92.5
CVD/RM	1	% of patients with a record of cholesterol measurement in the last 12 months	58.2	70.5	80.0
	2	% of patients with a record of smoking status	62.5	80.0	92.7
	3	% of patients with a record of a glucose measurement in the last 12 months	64.6	75.0	85.0
	4	% of patients with a record of blood pressure measurement in the last 12 months	80.0	87.5	92.5
	5	% of patients with a record of BMI*	27.5	47.6	70.0
	6	% of patients with a record of family history	17.5	33.7	57.0
	7	% of patients with a record of CVR◊ status	24.7	57.5	92.3
	8	% of patients with a record of a lipid lowering medication prescription	45.0	54.3	65.9
	9	% of patients with a record of a prescription of anticoagulants	39.0	51.2	77.5

** BMI = body mass index. ◊ CVR = cardiovascular risk.*

Analysis

We express the precision of a score as the 95% confidence limit either side of the mean, calculated as 1.96 times the SD[15]. For example, for a practice that scores 55% on the diabetes HbA1c indicator, a precision of 15 points indicates that we are 95% confident that the interval of 40% – 70% covers the 'true' score. We calculated the number of patient records needed to achieve a level of precision of 15, 10, and 5 points. The formulas used to calculate the sample size required at each level of precision are given in the Appendix. Inter-indicator Pearson correlations were calculated using SPSS 18.0.

To examine the impact on sample size of relaxing the criteria that performance scores at *all* practices attain the target level of precision, we calculated sample sizes required to achieve the target precision at 75% and 90% of all practices. This was to explore the benefit – in data collection terms – of reducing the sample by accepting less precise scores for a minority of practices.

To investigate the effect on precision of using fewer indicators, we began with a composite score made up of all indicators, and then removed indicators one-by-one. At each step, we removed the indicator that maximized the precision of a composite made from the remaining subset of indicators. This gave the 'optimal' subset of indicators: that is, the subset with the highest precision. We then repeated the process, but at each step removing the indicator that minimized the precision for the remaining indicators. This identified the least optimal indicator subsets (i.e. those with the lowest precision).

RESULTS

Precision of indicator performance scores

The sample sizes required to achieve different levels of precision were calculated for each practice separately, based on the practice's own observed scores. Table 2 summarizes the results for a selection of indicators that represent the range of findings. For each indicator, the first row shows the number of cases needed to achieve the target level of precision for *all* practices. The second and third rows show the sample sizes when a percentage of practices (10% and 25%, respectively) is allowed to have less precision than the target level. For example, a sample of 43 cases from each practice ensures that all practices will have their performance score on diabetes indicator 9 estimated with a precision of 15 points or better. This implies that at least 1 practice requires 43 cases, other practices may require fewer than 43 cases. As the degree of precision becomes greater (i.e. the confidence interval narrows), the number of cases needed increases: 384 cases per practice are required for all practices to achieve a precision of 5 points. This pattern of results is repeated for all indicators. However, when only 90% or 75% of practices are required to have scores within the target precision,

there is a substantial reduction in the sample size for some indicators, but very little change for others. A sample size of just 19 patients would yield a maximum confidence limit of 15 points on diabetes indicator 9 for 75% of practices. The corresponding sample size for CVD/RM indicator 7 is 42 patients, compared with 43 patients to reach this precision for all practices. Indicators where the more relaxed criteria resulted in the biggest reductions in sample size were those where most practices had either very high or very low performance scores.

Table 2. Number of cases required to achieve different levels of precision (95% confidence limit)

Condition	No.	% of practices achieving the target precision	No. of cases required for a target precision of:		
			15 points	10 points	5 points
Diabetes	3	All practices	43	96	384
		90% of practices	42	94	375
		75% of practices	36	81	323
	9	All practices	43	96	384
		90% of practices	27	60	239
		75% of practices	19	42	167
COPD	1	All practices	43	96	384
		90% of practices	42	95	380
		75% of practices	40	90	359
	3	All practices	43	96	384
		90% of practices	41	93	373
		75% of practices	37	84	335
CVD/RM	2	All practices	43	96	384
		90% of practices	42	95	380
		75% of practices	41	92	369
	7	All practices	43	96	384
		90% of practices	42	94	376
		75% of practices	35	79	314

Precision of composite performance scores

For each practice, a composite score was calculated for each of the three conditions. Practice mean composite scores and SDs appear in Table 3, as well as reliability coefficients corresponding to each level of precision. For diabetes and COPD, a precision of 15 points equates to a reliability coefficient of >0.7, the value generally taken to indicate acceptable reliability. However, CVD/RM scores demonstrate less variability between practices; here the same level of precision equates to a reliability coefficient of just 0.6. For all conditions, a precision of ≤10 points equates to reliability well above the threshold for acceptability.

Table 3. Mean composite scores and reliability coefficients corresponding to levels of precision

Condition	Indicator average Mean (SD)	Reliability coefficient at a level of precision of:		
		15 points	10 points	5 points
Diabetes	70.2 (15.4)	0.75	0.89	0.97
COPD	58.1 (14.7)	0.73	0.88	0.97
CVD/RM	48.8 (12.1)	0.60	0.82	0.96

Table 4. Number of cases required to achieve different levels of precision for composite performance scores

Condition	% of practices achieving the target precision	No. of cases required for a target precision of:		
		15 points	10 points	5 points
Diabetes	All practices	19	43	174
	90% of practices	9	20	81
	75% of practices	7	15	60
COPD	All practices	25	56	222
	90% of practices	17	38	152
	75% of practices	14	32	128
CVD/RM	All practices	40	89	356
	90% of practices	15	33	133
	75% of practices	11	25	101

For diabetes and COPD, the number of cases required to achieve a given precision on the composite score for *all* practices is around half the size of the sample required for the individual indicator scores (Table 4): for example 19 cases compared with 43 for a precision of 15 points for diabetes. However, for CVD/RM the reduction is only small (40 patients compared with 43). Even so, relaxing the target criteria to 90% or 75% of practices results in large reductions in sample size for all three conditions. The sample size gains from using a composite score are substantially greater for diabetes; this is related to the generally much higher practice performance scores.

Composite scores based on indicator subsets

Table 5 shows the sample sizes required to achieve a precision of 10 points, for differing subsets of indicators. For diabetes, an optimally selected subset of 7 indicators is sufficient to achieve a 10 point precision for all practices for a small increase in sample size (45 cases compared with 43 when using all 12 indicators). A sample size of just 20 can provide this level of precision for 90% of practices. However, other subsets of 7 diabetes indicators require a larger sample size to achieve the same precision, up to 34. Results are similar for CVD/RM, with subsets of 5 or 6 optimally selected indicators providing the same precision as the full indicator set for no appreciable increase in sample size.

Table 5. No. of cases required to achieve a precision of 10 points for subsets of composite scores (optimal selection - least optimal selection)

Condition	% of practices achieving the target precision	No. of indicators in composite score										
		12	11	10	9	8	7	6	5	4	3	2
Diabetes	All practices	43	44-43*	46-42*	46-45*	48-52	45-57	53-65	51-81	54-93	61-95	62-96
	90% of practices	20	20-22	19-24	20-26	19-29	20-34◊	21-40	21-46	22-52	22-59	28-73
	75% of practices	15	14-16	14-17	15-19	15-21	15-24	15-27	15-31	16-35	16-42	18-49
COPD	All practices	-	-	-	-	-	-	-	-	56	73-73	70-81
	90% of practices	-	-	-	-	-	-	-	-	38	41-49	49-64
	75% of practices	-	-	-	-	-	-	-	-	32	33-43	40-56
CVD/RM	All practices	-	-	-	89	89-90	90-94	88-94	93-94	92-94	92-94	91-95
	90% of practices	-	-	-	33	32-39	33-45	31-46	34-48	35-56	34-68	43-85
	75% of practices	-	-	-	25	25-30	25-35	24-37	26-41	27-47	27-60	34-81

Table 6. Mean composite performance scores and mean inter-indicator Pearson correlation for most optimal and least optimal subsets

Condition	Indicator subset	Practice mean	No. of indicators in composite score										
			12	11	10	9	8	7	6	5	4	3	2
Diabetes	Most optimal	Composite score	70.2	71.2	72.6	73.1	71.2	72.2	74.3	79.6	78.6	74.7	93.7
		Correlation	0.24	0.24	0.24	0.26	0.24	0.23	0.28	0.30	0.30	0.22	0.45
	Least optimal	Composite score	70.2	68.0	65.6	62.8	59.6	56.2	55.6	53.1	51.5	47.3	42.4
		Correlation	0.24	0.24	0.24	0.24	0.26	0.29	0.36	0.44	0.57	0.61	0.66
COPD	Most optimal	Composite score	-	-	-	-	-	-	-	-	58.1	53.2	60.3
		Correlation	-	-	-	-	-	-	-	-	0.21	0.17	0.08
	Least optimal	Composite score	-	-	-	-	-	-	-	-	58.1	51.5	55.8
		Correlation	-	-	-	-	-	-	-	-	0.21	0.26	0.38
CVRM	Most optimal	Composite score	-	-	-	48.8	47.5	45.0	47.3	47.7	44.5	47.8	32.7
		Correlation	-	-	-	0.23	0.21	0.20	0.17	0.18	0.16	0.06	-0.07
	Least optimal	Composite score	-	-	-	48.8	49.6	50.2	52.8	58.9	58.4	67.4	62.1
		Correlation	-	-	-	0.23	0.28	0.33	0.34	0.36	0.41	0.54	0.71

** For the subsets of 11, 10 and 9 diabetes indicators, the larger sample size for the most optimal subset is a consequence of the results for one practice. Without this practice, the number of cases required for the most optimal set were 36, 35 and 36 respectively. ◊ A composite score based on the 'most optimal' subset of 7 diabetes indicators has a precision of 10 points or less for 75% of the practices when based on a sample of 20 patients. For other subsets of 7 indicators, the required number of patients is higher, up to 34 patients.*

There are only 4 indicators in the full set for COPD; dropping any of these generally results in a bigger increase in sample size. When indicator subsets are selected in the least optimal manner, the impact varies more highly; in some scenarios (e.g. CVD/RM and 90% of practices) removal of just 2 indicators can result in a substantial increase in sample size requirement.
Table 6 shows the mean composite score and mean inter-indicator correlation across practices for each subset. The formula for the variance of the composite indicates that variance will be lower for subsets of indicators with smaller inter-indicator correlations and with scores closer to 0% and 100% (see Appendix). These characteristics are reflected in Table 6. For most subset sizes, the mean inter-indicator correlation is smaller for the most optimal, compared with least optimal, subset of indicators, with the difference increasing as the size of the subset decreases. The difference in mean composite score also increases as the subset size decreases.

DISCUSSION

The interpretation of performance scores requires caution when scores are based on small numbers of cases. In this study, for most individual indicators a sample of 96 cases was only just enough to compute a performance score with a 95% confidence limit of plus/minus 10 percentage points for many practices. By using composite scores, the same level of precision could be attained for the great majority of practices with samples a fraction of the size required for a single indicator. But even for composite scores, to achieve high levels of precision (i.e. confidence limits of ≤5 points) at all practices would still require the sampling of large numbers of patient records for some conditions. Low precision restricts the interpretability of the performance score and the ability to discriminate a performance score from a benchmark. However, collecting data on only a small number of indicators per condition does not necessarily lead to lower precision.

Strengths and limitations

We did not examine the impact on the composite score of weighting the indicators. A weighted score will have a precision somewhat greater or smaller than the unweighted composite, depending upon the variances of the indicators receiving the largest weights. However, if the weight assigned to an indicator is unrelated to performance on that indicator, then our results in effect represent a general average. Where weights are used to incentivize indicators with lower performance (as typically happens) and such indicators generally have higher variability, our results are likely to overestimate precision. We made some standard statistical assumptions that are detailed in the Appendix.

Implications

We found that the precision of a composite score is a function of both the performance scores on the individual indicators in the composite and the inter-correlations between those scores. Precision is greater for indicator sets displaying high or low performance scores, and low inter-indicator correlations. The effect of low inter-correlations might appear counterintuitive, but one way to understand the result is to recognize that a highly correlated indicator does not contribute very much additional unique information to the combined score. We would clearly not advocate selecting indicators on the basis of their high and low scores to maximize the precision of the composite: for assessment purposes it is important that indicators have 'room' to respond to changes in performance. However, where there is a choice among indicators with similar levels of performance, selecting a subset with generally smaller inter-correlations will have advantages in terms of precision.

Creators of quality indicator sets usually want the indicator performance scores to correlate, as an indication of coherence within the set of indicators, implying that the indicators are measuring one underlying concept[15,18]. For this reason one would want to be wary of including indicators with no, or only very weak, associations with the rest of the set, but equally the inclusion of highly correlated indicators may imply some redundancy. The mean inter-indicator correlations in this study were just over 0.2 for all the three conditions, this is one reason why our composites were so effective in reducing required sample sizes.

The precision of quality assessment is only weakly related to the number of indicators included in the composite score. Judicious selection of subsets of 5 to 7 diabetes or CVD/RM indicators generated composite scores with the same precision as the full indicator set without noticeably increasing the sample size requirement. However, by precision we mean the accuracy with which practice performance on this *specific set of indicators* is being measured. The incorporation of additional indicators may well be necessary to ensure the clinical and conceptual validity of the indicator set. In any debate about measuring a few conditions extensively or a broad range of conditions with only a few indicators per condition, our results would favor the latter.

In 2009/2010, the Quality and Outcomes Framework in the United Kingdom consisted of 86 clinical performance scores, with the number of indicators in each clinical area ranging from 2 for cancer and hypothyroidism to 18 for diabetes[1]. The HEDIS in the United States commonly uses only ≤5 indicators per condition[5]. Our results show that these smaller numbers of indicators per condition do not necessarily pose problems regarding precision. Incorporating more indicators might increase the validity of the instrument, but unless these are capturing new information they are unlikely to result in a notable benefit to the interpretability or discriminative ability of the scores.

The most effective way to increase the precision of both individual and composite performance scores is to increase the sample size. For smaller practices or conditions with low incidence and prevalence, larger sample sizes can be hard to establish. In the United States, United Kingdom, and the Netherlands, and indeed throughout Europe, there are many solo or small practices. There are 2,340 people in the Dutch population per fulltime-equivalent physician[21]. This suggests that for solo practices, reasonably precise performance scores (i.e. with a 95% confidence limit of 10 points) for individual indicators can only be derived for conditions with a prevalence of >4%. Using composite scores it should be possible to reach the same precision for conditions with a prevalence of 2% depending upon the condition. Even so, there will be many conditions for which precise measurement is not feasible at smaller practices.

Conclusions

On the basis our findings, policy makers should be cautious when interpreting performance scores for individual indicators based on small numbers of cases. A well-designed composite measure can provide a more precise measure of performance; a small number of indicators included can be sufficient. Data collection demands can be substantially reduced by using a sample size that accepts that the precision of measurement will be below target for a small minority of practices.

When selecting indicators it is important to find a good balance between considerations of the validity of the composite measure – for which good correlations between indicators might be one criterion – and the precision of the score – for which low inter-correlations are desirable. Other aspects that should be taken into account include the desired level of comprehensiveness and the feasibility of data extraction[24]. It is crucial to align issues for assessment to professional priorities and to issues of clinical relevance to patients and policy makers, as this is more likely to influence and change professional behavior[25]. The degree of precision required is also related to the purpose of the assessment; in particular whether indicators are being used for external summative judgment or formative internal quality improvement[18,12]. Precision needs to be greatest when there is an external summative assessment and where assessment is used as a payment mechanism. Different situations (summative, formative, local, national, etc.) may require a different focus. Our results show that especially for formative quality improvement, composite scores based on fewer cases and a small number of (low-correlated) indicators can provide a precise enough score.

REFERENCES

1. website Quality and Outcomes Framework for 2012/2013, guidance for PCOs and practices, NHS Employers 2012. http://www.nhsemployers.org/Aboutus/Publications/Documents/QOF_2012-13.pdf.
2. Campbell SM, Roland MO, Quayle JA, Buetow SA, Shekelle PG. Quality indicators for general practice: which ones can general practitioners and health authority managers agree are important and how useful are they? J Pub Health Med 1998;20(4):414-21.
3. Roland M. Linking physicians' pay to the quality of care: a major experiment in the UK. N Engl J Med 2004;351(14):1448-54.
4. website National Committee For Quality Assurance, HEDIS and Quality Measurement. http://www.ncqa.org/HEDISQualityMeasurement.aspx.
5. website National Committee For Quality Assurance, HEDIS 2012. http://www.ncqa.org/HEDISQualityMeasurement/HEDISMeasures/HEDIS2012.aspx.
6. Burge FI, Bower K, Putnam W, Cox JL. Quality indicators for cardiovascular primary care. Can J Cardiol 2007;23(5):383-8.
7. Gardner KL, Sibthorpe B, Longstaff D. National quality and performance system for Divisions of General Practice: early reflections on a system under development. Aust New Zealand Health Policy 2008;5:8.
8. Braspenning J, Dijkstra R, Tacken M, Bouma M, Witmer H. Visitatie Instrument Accreditering (VIA®). Nijmegen/Utrecht: Kwaliteit van Zorg (WOK); UMC St Radboud; Nederlands Huisartsen Genootschap (NHG); NHG Praktijk Accreditering BV (NPA); 2007.
9. van den Hombergh P, Grol R, van den Hoogen HJ, van den Bosch WJ. Assessment of management in general practice: validation of a practice visit method. Br J Gen Pract 1998; 48(436):1743-50.
10. Sequist TD, Schneider EC, Li A, Rogers WH, Safran DG. Reliability of medical group and physician performance measurement in the primary care setting. Med Care 2011;49(2):126-31.
11. Lyratzopoulos G, Elliott MN, Barbiere JM, Staetsky L, Paddison CA, et al. How can health care organizations be reliably compared? Lessons from a national survey of patient experience. Med Care 2011;49(8):724-33.
12. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. BMJ 2003;326(7393):816-9.
13. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2003 (3rd ed.).
14. Campbell SM, Kontopantelis E, Hannon K, Burke M, Barber A, et al. Framework and indicator testing protocol for developing and piloting quality indicators for the UK quality and outcomes framework. BMC Fam Pract 2011;12:85.
15. Bernstein S, Bernstein R. Schaum's outline of elements of statistics 2: inferential statistics. New York, NY: McGraw-Hill; 1999.
16. Nunnally JC, Bernstein IH. Psychometric Theory. New York, NY: McGraw-Hill; 1994 (3rd ed.).
17. Raleigh VS, Foot C. Getting the measure of quality: opportunities and challenges. London: Kings Fund; 2010.
18. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, et al. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med Care 2007; 45(6):489-96.
19. Scholle SH, Roski J, Adams JL, Dunn DL, Kerr EA, et al. Benchmarking physician performance: reliability of individual and composite measures. Am J Manag Care 2008;14(12):833-8.
20. Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Improving the reliability of physician performance assessment: identifying the "physician effect" on quality and creating composite measures. Med Care 2009; 47(4):378-87.
21. Hingstman L, Kenens RJ. Cijfers uit de registratie van huisartsen – peiling 2008. Utrecht: NIVEL; 2009.
22. Mant J. Process versus outcome indicators in the assessment of quality of health care. Int J Qual Health Care 2001;13(6):475-80.
23. Braspenning JCC, Pijnenborg L, In 't Veld CJ, Grol RPTM (redactie). Werken aan kwaliteit in

de huisartsenpraktijk: indicatoren gebaseerd op de NHG-Standaarden. Houten: Bohn Stafleu van Loghum; 2005.
24. Campbell SM, Scott A, Parker RM, Naccarella L, Furler JS, et al. Implementing pay-for-performance in Australian primary care: lessons learned from the United Kingdom and the United States. Med J Aust 2010;193(7):408-11.
25. McDonald R, Roland M. Pay for performance in primary care in England and California: comparison of unintended consequences. Ann Fam Med 2009;7(2):121-7.

REFERENCES APPENDIX

1. Hoel PG. Introduction to mathematical statistics. Toronto, Ontario: John Wiley & Sons Canada; 1971 (4th ed.): 82.
2. Box GE, Hunter WG, Hunter JS. Statistics for experimenters: an introduction to design, data analysis, and model building. University Michigan: Wiley; 1978: 130.

APPENDIX

To compute sample sizes for the present paper, we utilized a number of formulas, described below. While the methods used for individual indicators are quite standard, those for composite indicators were derived specifically for this paper.

Precision of a performance score for a single indicator

For a given practice, let p_i be the observed performance score on indicator i, expressed as a proportion (i.e. $0 \leq p_i \leq 1$), and let n be the number of patients (observations) on which p_i is based. Then, using the standard Normal approximation to the binomial distribution[1], the variance of p_i can be estimated in the usual way as:

$$V(p_i) = \frac{p_i(1-p_i)}{n} \qquad (1)$$

The distribution of p_i is reasonably Normal for $np_i \geq 5$ (Box et al.[2], p. 130).

Let C be the desired level of precision (confidence limit) on p_i, also as a proportion. Then, the number of cases n required to achieve this level of precision is given by:

$$n = \frac{Z^2 p_i(1-p_i)}{C^2} \qquad (2)$$

Where Z is the standardized Normal variate associated with the required confidence limit. For example, $Z = 1.96$ for a 95% confidence limit.

Precision of a composite performance score

Let $p_1, p_2, \ldots p_k$ be the observed practice performance scores on a set of k indicators, and let the composite score, p_c, derived from these be the mean of the k scores:

$$p_c = \frac{\sum_{i=1}^{k} p_i}{k} \qquad (3)$$

The variance of p_c is given by:

$$V(p_c) = \frac{1}{k^2} V\left(\sum_{i=1}^{k} p_i\right) \qquad (4)$$

Using standard distributional theory and the Normal approximation to the binomial the variance of p_c can be estimated as:

$$V(p_c) = \frac{1}{k^2}\left(\sum_{i=1}^{k} \frac{p_i(1-p_i)}{n_i} + 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} r_{ij}\sqrt{\frac{p_i(1-p_i)p_j(1-p_j)}{n_i n_j}} \right) \quad (5)$$

Where r_{ij} = observed correlation between indicators i and j.

In the special case when n_i has the same value, n, for all k indicators, such as with the present dataset:

$$V(p_c) = \frac{1}{nk^2}\left(\sum_{i=1}^{k} p_i(1-p_i) + 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} r_{ij}\sqrt{p_i(1-p_i)p_j(1-p_j)} \right) \quad (6)$$

Where r_{ij} = observed correlation between indicators i and j.

The sample size required to achieve a precision of size C on the composite score is then given by:

$$n = \frac{1}{k^2C^2}\left(Z^2\left(\sum_{i=1}^{k} p_i(1-p_i) + 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} r_{ij}\sqrt{p_i(1-p_i)p_j(1-p_j)} \right) \right) \quad (7)$$

We note from equations (6) and (7) that $V(p_c)$ and n are reduced in size when (I) performance scores are further away from a value of 0.5, and (II) indicator inter-correlations are small. Thus for a given sample size, precision will be higher (i.e. the confidence interval narrower) for a set of indicators with high and/or low performance scores and low inter-correlations compared to a same-sized set with middling performance scores and larger inter-correlations.

Calculation of reliability

We let $V_T(p_c)$ represent the variance between practices in 'true' composite performance scores and $V_O(p_c)$ the variance in observed practice scores, the coefficient of reliability is then defined in the standard way as the ratio of true score variance to observed score variance:

$$r = \frac{V_T(p_c)}{V_O(p_c)} \quad (8)$$

We previously defined $V(p_c)$ to be the variance of the observed composite score for an individual practice. Assuming that $V(p_c)$ is the same for all practices, the variance in observed scores across practices is equal to the true variance plus the within-practice variance:

$$V_O(p_c) = V_T(p_c) + V(p_c) \quad (9)$$

From equations (8) and (9), reliability can be estimated as:

$$r = \frac{V_o(p_c) - V(p_c)}{V_o(p_c)} \qquad (10)$$

The variance around a practice score can be expressed in terms of the associated precision, C, as:

$$V(p_c) = \frac{C^2}{Z^2} \qquad (11)$$

Hence the reliability coefficient can be estimated from the precision and the observed between-practice variance using:

$$r = \frac{V_o(p_c) - \frac{C^2}{Z^2}}{V_o(p_c)} \qquad (12)$$

We used formula (12) to estimate reliability values for the paper. The formula rests on the assumption that $V(p_c)$ is the same for all practices. In practice that will not be so, in which case r is the reliability corresponding to the desired precision C.

Identification of reduced sets of indicators

We investigated the effect on precision of using fewer indicators in the composite, by first constructing the composite using all indicators and then removing indicators one-by-one, at each step removing the indicator that minimized the variance of the composite score for the remaining subset of indicators. To do this, on the basis of equation (6) above, at each step we computed Y_i below for each indicator i.

$$Y_i = p_i(1 - p_i) + 2\sqrt{p_i(1 - p_i)}\sum_{j \neq i}^{m} r_{ij}\sqrt{p_j(1 - p_j)} \qquad (13)$$

Where m is the number of indicators in the composite at each step.

We computed a separate Y_i for each practice and summed these across practices. We then identified the indicator with the largest sum. Removal of this indicator gave the reduced set of indicators with the minimum total sum of within-practice variances for the resulting composite score.

This process identified the 'optimal' subset of indicators at each step: that is, the subset with the smallest possible sum of within-practice variances. We then repeated the process, again starting with the full set of indicators, but at each step removing the indicator that *maximized* the sum of practice variances for a composite of the remaining indicators. This identified the least optimal indicator subset at each step (i.e. the one with the maximum variance). We used

the results to compute the sample sizes required for composites based on the optimal and least-optimal indicator subsets at each step to have a precision of 10 percentage points.

Finite population adjustment

For maximum generalizability, the equations for sample size above assume that the patients measured on an indicator at a practice are a random selection from a very much larger – effectively infinite – population of potential patients. However, a 'finite population' adjustment would be appropriate if the goal were to measure performance solely for a fixed set of patients who attend the practice. In this case, we regard the patients eligible for an indicator at a particular practice as constituting the total population of eligible patients, such that if all these patients were assessed the error on the performance score would by definition be zero (i.e. perfect precision). For this purpose, the sample sizes reported in the paper can be adjusted using the below formula:

$$n_f = \frac{nN}{n+N-1} \qquad (14)$$

Where n is the required sample assuming an infinite population, N is the number of patients in the finite population, and n_f is the corresponding sample required from this population.

As an example, our paper reports that a sample of 90 patients is required to achieve a precision of 10 percentage points for the diabetic care composite score. For a practice with 150 diabetic patients, the number of these patients that has to be sampled to estimate the performance score for the full group of 150 to a precision of 10 points is:

$$n_f = \frac{90 \times 150}{90+150-1} = 56$$

Excluded data

For each of the conditions a small number of patients were excluded due to a completely empty medical record. We excluded substantial numbers of diabetes and COPD patients (20.2% and 35.1% respectively) where the main care provider was not the FP but a specialist. As a final step practices that, after the above exclusions, had data for less than 20 patients with a condition were excluded from the dataset for that condition. We imposed this step to reduce the occurrence of poorly estimated indicator performance scores, particularly spurious 0% and 100% scores that bias variance estimates (see below). For a large number of COPD patients the FP was not the main care provider, resulting in a larger percentage of both patients and practices being excluded for this condition. Table 1 summarizes the included and excluded data.

Table 1. Amount of included and excluded data

Condition	Total amount of data		Included in calculations		% excluded	
	practices	*patients*	*practices*	*patients*	*practices*	*patients*
Diabetes	392	15033	350	11783	10.7	21.6
COPD	455	15290	286	8128	37.1	46.8
CVD/RM	379	14267	342	14158	9.8	0.8

Treatment of performance scores of 0% and 100%

Some practices had a performance score of 0 or 1 on some indicators. When used with equation (1), these scores return a variance estimate of zero. Since zero variance is unrealistic, we replaced 0% and 100% scores with a score close to 0 or 100 using the following formulas:

When $p_i = 0$ we substituted $p_i = \frac{d}{n}$

When $p_i = 1$ we substituted $p_i = \frac{n - d}{n}$

Where $d = 0{,}5$ and n = number of cases for the practice.

For 5 of the 12 diabetes indicators we applied this to more than 10% of practice scores. This was especially the case with diabetes indicator 3, where 94 practices (26%) scored 0%. The CVD/RM set also included 3 indicators (indicators 5, 6 and 7) with a large number of 0% or 100% scores. For COPD, only a small percentage (less than 5 %) of scores were amended.

Statistical assumptions

The underlying data has a binomial distribution to which we have applied the Normal approximation. Using standard criteria related to sample size and score[2], the approximation is good for the majority of cases we have examined, the main exceptions being for the lowest sample size (20 cases) coupled with a performance scores below 25% or above 75%, in which case our estimates of precision do not represent the skew in the actual distribution. For samples of 40 or more, the approximation is good for performance scores between 12.5 and 88.5.

Our approach to estimating precision, reliability, and required sample sizes followed the standard statistical methodology of substituting observed practice indicator scores, variances, and inter-indicator correlations for the unobserved 'true' population scores. Any error in the overall results resulting from this will be small however, as over-estimated and under-estimated values should be reasonably balanced across practices.

4

Effect of accreditation on the quality of chronic disease management: a comparative observational study

Arna van Doorn - Klomberg
Jozé Braspenning
René Wolters
Margriet Bouma
Michel Wensing

BMC Family Practice 2014; 15: 179.

ABSTRACT

Background Practice accreditation is widely used to assess and improve quality of healthcare providers. Little is known about its effectiveness, particularly in primary care. In this study we examined the effect of accreditation on quality of care regarding diabetes, chronic obstructive pulmonary disease (COPD) and cardiovascular disease (CVD).

Methods A comparative observational study with two cohorts was performed. We included 138 Dutch general practices that participated in the Dutch Accreditation program for primary care. A first cohort of 69 practices was measured at start and completion of the three-year Accreditation program. A second cohort of 69 practices was included and measured simultaneously with the final measurement of the first cohort. In separate multilevel regression analyses, we compared both within-group changes in the first cohort and between-groups differences at follow-up (first cohort) and start (second cohort). Outcome measures consisted of 24 systematically developed indicators of quality of care in targeted chronic conditions.

Results In the within-group comparison, we found improvements on 6 indicators related to diabetes (feet examination, cholesterol measurement, lipid lowering medication prescription) and COPD (spirometry measurement, smoking cessation advice). In the between-groups comparison we found that first cohort practices performed better on 4 indicators related to diabetes (cholesterol outcome) and CVD (blood pressure outcome, smoke status registration, glucose measurement).

Conclusions Improvements of the quality of primary care for patients with chronic conditions were found, but few could be attributed to the Dutch Accreditation program. Further development of accreditation is needed to enhance its effectiveness on chronic disease management.

BACKGROUND

A range of strategies has been developed to assess and improve healthcare for patients with chronic conditions[1]. A widely used strategy is accreditation of healthcare providers[2,3]. Accreditation affects the institution or practice and is offered more or less voluntary, while certification focuses on a specific norm that should be reached by individuals or particular services[4,5]. A key element of accreditation is audit and feedback. A Cochrane review suggested that an audit and feedback system has a small positive effect on quality of care overall[6], but the added value of accreditation was not considered. Greenfield et al.[7] suggested in their systematic review that accreditation can promote change, for example through the opportunity to reflect on organizational performance, and that it has an effect on professional development. However, they reported inconsistent findings regarding quantifiable effects of accreditation on measures of clinical processes and outcomes. Few rigorous evaluations of effectiveness of accreditation are available, particularly in primary care[8-10]. O'Beirne et al.[10] performed a review of peer-reviewed and grey literature regarding accreditation in primary care. They found indications that accreditation may improve both organization and outcomes of care, but conclude that more research is needed. Szecsenyi et al.[11] examined an accreditation program that focuses on practice management; they found improvements on several quality and safety measures regarding complaint management, analysis of critical incidents and quality development. Given the shortage of controlled evaluations, more research on the effects of well-defined accreditation programs is required.

In this study we assessed an accreditation program for primary care practices[12,13], set up by the Dutch College of General Practitioners in 2005. This program strongly focuses on the educational value of audit and feedback and participation is voluntary. However, part of the program does require meeting specific norms, for example on hygiene, first aid equipment and accessibility. The program addresses clinical care as well as practice management. The clinical care section consists mainly of three chronic care conditions: diabetes mellitus, chronic obstructive pulmonary disease (COPD) and cardiovascular disease (CVD).

In the same period, health insurers developed bundled payment schemes, with an initial focus on diabetes (approximate start in 2008), and in later years also on COPD (since 2009) and CVD (since 2011)[14,15]. Bundled payment is suggested as a method to control costs while providing more patient-centered, higher quality care[16]. An international review of different types of bundled payment[17] suggests that there is weak but consistent evidence for a positive effect of bundled payment on healthcare costs, with no tangible impact on quality of care. Studies on the effect of bundled payment in the Netherlands show that it improved organization and management of care[18,19] and potentially also the outcomes of diabetes

care[20]. Nevertheless, a Cochrane review concluded that there is insufficient evidence to draw firm conclusions on the effectiveness of different these types of reimbursement schemes[21].
This study is focused on effects of the Dutch Accreditation program in the context of a healthcare system, in which provider reimbursement of chronic care was improved simultaneously. In the Netherlands, approximately 80% of diabetes patients has the family physician (FP) as the principal caregiver. This percentage is slightly lower for COPD and CVD patients (50-65%). Generally, patients with the FP as main care provider tend to be stable, although not necessarily well controlled. The study aimed to determine whether participation in the Accreditation program led to improvements in quality of care regarding diabetes, COPD and CVD. We also aimed to assess whether the improved performance was different from the performance during pre-assessment of a second group of participants.

METHODS

Study design

We performed a comparative observational study with two cohorts of primary care practices. A first cohort of practices was followed throughout the three-year program with extended measurements at start and after three years. A second cohort was measured at their start of the Accreditation program, during the same period of the post-assessment in the first cohort. Data collected at the start of the practice accreditation process portrayed practice performance prior to the accreditation process: data were collected retrospectively over a one-year period prior to the start. We evaluated the effect of practice accreditation in a design with two separate comparisons. In the first comparison we assessed improvements over time within the first cohort. In the second comparison we assessed differences between the first and second cohort by comparing the follow-up measurement of the first cohort with the start measurement of the second cohort. Since both of these groups collected data in the same period, this comparison could offer additional information on effects of accreditation (3 years experience versus start of accreditation). The ethics committee of the Radboud University Medical Center provided a waiver for the study.

Study population

Practices included in the study were Dutch primary care practices that participated in the Accreditation program of the Dutch College of General Practitioners between 2006 and 2011. We included a group of 69 practices in the first cohort that joined the Accreditation program before 2009 and had collected follow-up data. Start measurements took place between 2006 and 2008; follow-up measurements of this cohort took place between 2009 and 2011.

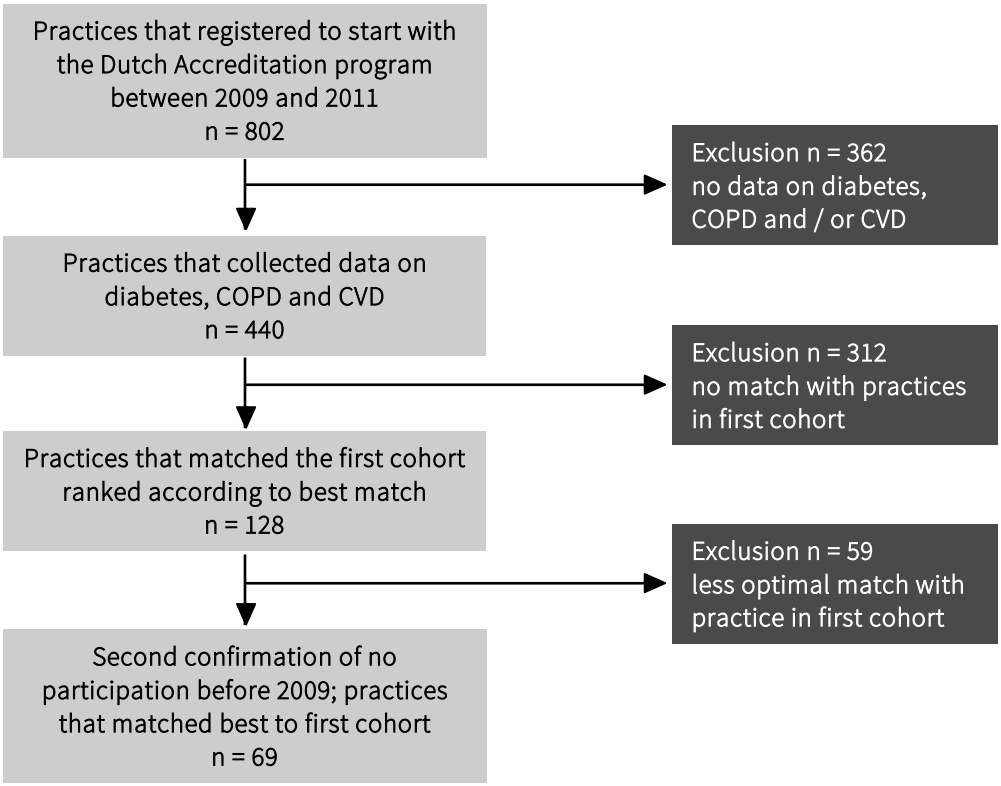


Figure 1. Selection of practices in the second cohort

The average time period between data collection in the first cohort was 4 years and 2 months, with a minimal interval of 3 years and 2 months. The second cohort consisted of primary care practices that had their start measurement between 2009 and 2011. We used a matched design to select an equally large sample of 69 practices (see Figure 1). Matching was based on availability of data, practice location (degree of urbanization), visitation date, practice type, and practice size.

Practice accreditation

Practices have been invited to participate voluntarily in the Dutch Accreditation program since 2005. The preparatory phase of the program consisted of data collection regarding practice management and patient care. The measurement instruments used were previously validated questionnaires such as the 'VIP', a visitation instrument for practice organization[22] and the 'Europep' for patient experiences with care[23]. The clinical performance indicators were derived from the national guidelines valid at the time of measurement[24]. These indicators were measured with the use of patient information that was extracted from electronic medical records; the family physician (FP) or nurse extracted the information either automatically or manually with a standardized extraction form. After data submission via an

online system, the practice received a report that included information on its own performance and the performance of other practices as a benchmark. This information could be used to identify which areas needed improvement. The FPs wrote improvement plans with a plan-do-study-act cycle. The first audit was carried out after the approval of these plans to confirm adequate participation and to grant accreditation. After this audit a three-year accreditation cycle started. At the end of each year the practice staff evaluated whether the objectives of improvement plans were met and wrote new improvement plans for the following year. The prolongation of accreditation depended on this process; accreditation was not based on the actual quality of care itself, but rather on the quality of the improvement initiatives according to a structured program.

Measures

The primary care practice maintains comprehensive patient records, which includes information on all contacts and procedures in primary care, hospital, and several other care providers. For this study, we excluded all patients of which the FP was not the main care provider. All information used in this study was collected as part of the Accreditation program. The prevalence of diabetes and COPD was defined as the number of patients with the condition divided by the total number of patients in the practice. Indicators of both processes of care and intermediate outcomes were included regarding patients with diabetes, COPD and CVD (Table 1). The data that were collected before 2009 on CVD related to a broader range of patients; patients with high risk for CVD were also included. Since this made it impossible to compare data over time we only included data after 2009, that relate to patients with known CVD. During our research period several target levels as described in the guidelines shifted for diabetes (relating to indicators 2 and 3) and for CVD (indicator 2). We calculated the percentage of patients with a level below target level according to the target level that was valid at the time of measurement. Data on practice characteristics were collected through the use of online questionnaires which were also part of the Dutch Accreditation program.

Analysis

Descriptive statistics were used to describe the practice population and gain insight in the performance scores. For each of the three conditions we calculated a mean score for several process indicators. Some indicators could not be included in the mean scores due to low internal consistency with other indicators or use of subsamples: diabetes indicator 5, COPD indicator 3 and CVD indicators 4, 5 and 6 (Table 1).

Table 1. Overview of indicators

Condition	No.	Description
Diabetes		
Outcome	1	% of patients with an HbA1c value within target level (53 mmol/mol or less)
indicators	2	% of patients with a blood pressure value within target level (*2006-2008*: 150/80 mm Hg or less, *2009-2011*: systolic 140 mm Hg or less)
	3	% of patients with a total cholesterol value within target level (*2006-2008*: 5.0 mmol/l or less, *2009-2011*: 4.5 mmol/l or less)
Process	4	Calculated mean process score based on the following 6 performance measures:
indicators	4a	% of patients with a record of HbA1c measurement in the last 12 months
	4b	% of patients with a record of blood pressure measurement in the last 12 months
	4c	% of patients with a record of cholesterol measurement in the last 12 months
	4d	% of patients with a record of creatinine measurement in the last 12 months
	4e	% of patients with a record of retinal examination in the last 24 months
	4f	% of patients with a record of feet examination in the last 12 months
	5	% of patients with a current record of a lipid lowering medication prescription
COPD		
Outcome	1	% of patients that smoke
Process	2	Calculated mean process score based on the following 2 performance measures:
indicators	2a	% of patients with a record of smoking status
	2b	% of patients with a record of a spirometry assessment in the last 12 months
	3	% of patients who smoke with a record of smoking cessation advice
CVD		
Outcome	1	% of patients that smoke
indicators	2	% of patients with a blood pressure value within target level (*2006-2008*: 160/90 mm Hg or less, *2009-2011*: systolic 140 mm Hg or less)
Process	3	Calculated mean process score based on the following 2 performance measures:
indicators	3a	% of patients with a record of smoking status
	3b	% of patients with a record of blood pressure measurement in the last 12 months
	4	% of patients who smoke with a record of smoking cessation advice
	5	% of patients with a record of a glucose measurement in the last 12 months
	6	% of patients with a current record of a prescription of anticoagulants

In order to examine whether chronic care management improved after the three-year cycle of the Accreditation program, we compared the start and follow-up scores of the first cohort in a within-subject design. We performed separate multilevel regression analyses for each indicator. We accounted for repeated measures per practice and included predictors reflecting practice size, practice type and practice location. Practices in the first cohort were compared to practices in the second cohort on follow-up and baseline scores respectively, in a separate regression analysis. The comparison was corrected for the match criteria practice size, practice type and practice location in each model. All analyses were performed with SPSS version 20[25].

Table 2. Practice characteristics of the first and second cohort

Practice characteristics	Practices in first cohort 2006-2008		2009-2011		Practices in second cohort, 2009-2011	
	n	*%*	*n*	*%*	*n*	*%*
Practice type						
Single handed	17	25%	18	26%	19	27%
Health center	10	14%	11	16%	8	12%
Other	42	61%	40	58%	42	61%
*Practice location**						
Urban	23	33%	23	33%	25	36%
Partly urban	34	49%	34	49%	33	48%
Rural	12	18%	12	18%	11	16%
Average no. of patients per practice	4629		4808		4830	

** A region was defined as urban when the number of addresses per km^2 exceeded 1,500; partly urban regions had 500 - 1.500 addresses per km^2; rural regions had less than 500 addresses per km^2.*

RESULTS

Study population

Table 2 shows the percentages and means of different practice characteristics in each group. There was a slight underrepresentation of practices that were part of a health center in the second cohort compared to the first cohort: 8 (12%) versus 11 (16%) practices. They were slightly more often located in an urban area (25 (36%) versus 23 (33%) practices). The mean number of patients in the practice increased in time; mean practice size was similar in the two cohorts.

Within-group comparison

The first two columns in Table 3 report on scores at start and follow-up of practices in the first cohort. The recorded prevalence of diabetes increased from 40 to 50 per 1,000 patients. The recorded prevalence of COPD also increased, from 17 to 22 per 1,000 patients. The prevalence of CVD/CVRM was not measured. Table 3 also reports results from the multilevel regression analyses. Regarding diabetes we found that practices had improved on three performance indicators. Practice scores had higher percentages of patients with a recorded cholesterol measurement ($p = 0.04$), feet examination ($p = 0.03$) and prescription of lipid lowering medication ($p = 0.009$). The percentage of patients with a blood pressure below target level was marginally lower at follow up ($p = 0.08$). Practices in the first cohort had improved on 3 COPD process indicators. There was an increase in the percentage of patients with a recorded spirometry measurement ($p = 0.000$), which was the sole contributor to the effect on the mean process score ($p = 0.01$). We also found a large improvement in the percentage of patients who smoke that received an advice to stop smoking ($p = 0.002$).

Table 3. Within-practice comparison of first and second measurement of practices in first cohort and between practice comparison of first and second cohort in 2009-2011

Indicators*	1st cohort 2006-2008 *Mean (SD)*	2009-2011 *Mean (SD)*	Within-group comparison of 1st cohort◊ *Estimate (95% CI)*	*p*	2nd cohort 2009-2011 *Mean (SD)*	Between-groups comparison 1st and 2nd cohort 2009-2011‡ *Estimate (95% CI)*	*p*
Diabetes							
Prevalence	0.040 (0.014)	0.050 (0.013)	0.010 (0.006 – 0.015)	0.000	0.051 (0.018)	-0.001 (-0.006 – 0.005)	0.79
Outcome measures							
HbA1c ≤ target	60.8 (15.9)	63.6 (11.4)	2.6 (-2.6 – 7.9)	0.32	63.6 (11.0)	-0.02 (-4.0 – 3.9)	0.99
Blood pressure ≤ target^	56.6 (16.7)	52.1 (17.0)	-5.4 (-11.4 – 0.7)	0.08	50.8 (16.7)	0.9 (-4.9 – 6.7)	0.76
Total cholesterol ≤ target^	50.4 (15.0)	45.2 (20.1)	-5.4 (-12.4 – 1.5)	0.12	36.3 (20.7)	8.8 (1.5 – 16.1)	0.02
Process measures							
Mean process score	79.2 (10.0)	81.6 (12.7)	2.3 (-1.5 – 6.1)	0.23	79.5 (12.7)	1.8 (-2.5 – 6.2)	0.41
HbA1c measurement	90.4 (10.2)	92.5 (7.3)	2.1 (-1.1 – 5.4)	0.20	92.4 (6.7)	-0.03 (-2.5 – 2.4)	0.98
BP measurement#	93.7 (7.0)	91.9 (13.3)	-1.7 (-5.0 – 1.7)	0.32	92.4 (8.9)	-0.7 (-4.7 – 3.3)	0.74
Cholesterol measurement	81.3 (14.4)	86.1 (12.4)	4.8 (0.2 – 9.4)	0.04	86.6 (9.4)	-0.6 (-4.4 – 3.2)	0.74
Creatinine measurement	82.1 (13.1)	85.3 (17.0)	3.4 (-1.2 – 7.9)	0.14	83.6 (17.3)	1.2 (-4.8 – 7.2)	0.69
Retinal examination	70.6 (16.4)	66.4 (24.4)	-4.5 (-12.0 – 3.1)	0.24	62.3 (26.6)	3.9 (-4.9 – 12.7)	0.39
Feet examination	57.3 (22.7)	67.1 (26.4)	9.7 (0.9 – 18.5)	0.03	59.7 (27.1)	7.2 (-2.1 – 16.5)	0.13
Lipid lowering medication prescription	60.5 (16.4)	69.9 (22.6)	9.5 (2.5 – 16.4)	0.009	68.4 (13.1)	1.1 (-5.1 – 7.4)	0.72
COPD							
Prevalence	0.017 (0.009)	0.022 (0.008)	0.005 (0.003 – 0.008)	0.000	0.023 (0.010)	-0.001 (-0.004 – 0.003)	0.65
Outcome measure							
Patients that smoke	36.6 (22.9)	31.8 (16.1)	-4.9 (-11.5 – 1.8)	0.15	32.2 (20.7)	-0.4 (-6.9 – 6.2)	0.92
Process measures							
Mean process score	58.5 (17.8)	67.2 (22.1)	8.3 (1.8 – 14.8)	0.01	61.6 (24.9)	5.5 (-2.8 – 13.8)	0.19
Record of smoke status	76.0 (21.4)	75.4 (23.2)	-1.3 (-8.1 – 5.6)	0.71	69.4 (27.0)	6.1 (-2.8 – 15.0)	0.18
Spirometry measurement	41.2 (23.5)	58.8 (28.7)	17.0 (8.0 – 26.0)	0.000	53.8 (28.5)	4.7 (-5.5 – 14.8)	0.37
Smoking cessation advice	47.0 (33.9)	69.8 (30.8)	21.8 (8.7 – 34.9)	0.002	65.2 (32.6)	5.1 (-10.1 – 20.2)	0.51

Indicators*	1st cohort 2006-2008	2009-2011	Within-group comparison of 1st cohort◊		2nd cohort 2009-2011	Between-groups comparison 1st and 2nd cohort 2009-2011‡	
	Mean (SD)	*Mean (SD)*	*Estimate (95% CI)*	*p*	*Mean (SD)*	*Estimate (95% CI)*	*p*
Cardiovascular disease†							
Outcome measures							
Patients that smoke	-	12.6 (8.5)	-	-	10.5 (7.8)	1.9 (-1.1 – 4.9)	0.20
Blood pressure ≤ target^	-	41.8 (15.5)	-	-	35.7 (14.7)	6.3 (0.6 – 11.9)	0.03
Process measures							
Mean process score	-	61.3 (20.3)	-	-	54.5 (21.6)	6.5 (-1.2 – 14.1)	0.10
Record of smoke status	-	51.7 (26.6)	-	-	39.8 (25.5)	11.3 (1.9 – 20.8)	0.02
BP measurement#	-	70.8 (17.9)	-	-	69.3 (21.3)	1.6 (-5.6 – 8.9)	0.66
Smoking cessation advice	-	66.7 (34.2)	-	-	51.1 (34.0)	13.2 (-4.6 – 30.9)	0.14
Glucose measurement	-	77.5 (17.6)	-	-	64.3 (24.3)	12.0 (1.9 – 22.0)	0.02
Prescription of anticoagulants	-	78.8 (15.0)	-	-	76.5 (20.4)	1.6 (-5.1 – 8.3)	0.64

** Except for prevalence values, all values in this table report the percentage of patients treated according to the guidelines / for which the indicator was met (deviations, estimates and 95% confidence intervals are also reported in number of percentage points). ◊ Results are based on multilevel regression analyses, accounting for repeated measures per practice and correcting for practice size, practice type and practice location. ‡ Results are based on regression analyses, correcting for practice size, practice type and practice location. ^ In prevailing guidelines, target levels on blood pressure and cholesterol were tightened over time. This may partly account for a decrease in these scores regarding diabetes. # BP = blood pressure. † The inclusion criteria for patients with risk for cardiovascular disease changed towards the inclusion of patients with known cardiovascular disease only, which made a within-group comparison of the first cohort not justifiable.*

Between-groups comparison

The figures for the between-groups comparison are reported in the columns on the right side of Table 3. No differences were found in the recorded prevalence of diabetes and COPD. Diabetes patients in the second cohort less often had a total cholesterol value within target level compared to the patients in the first cohort (p = 0.02). No other differences were found between the two cohorts regarding diabetes. We did not find any differences on COPD performance indicators. Practices in the first cohort provided higher quality of care than the second cohort on 3 CVD performance indicators. A higher percentage of patients had a blood pressure within target level (p = 0.03). Furthermore, the number of patients with recorded smoke status (p = 0.02) and glucose values (p = 0.02) had increased.

DISCUSSION

Improvements were found regarding cholesterol measurement, feet examination and the prescription of lipid lowering medication (diabetes), spirometry measurement and the provision of a smoking cessation advice (COPD). As the follow-up measurement scores on these performance measures were similar to the baseline scores in the second cohort, it remains uncertain whether the effect can be attributed to the Dutch Accreditation program. Compared to the second cohort of newly starting practices, the practices in the first cohort performed better regarding achievement of a target cholesterol level (diabetes), blood pressure level (CVD) and the registration of smoke status and measurement of glucose levels (CVD). These differences could provide evidence for the added value of the Accreditation program.

Explanation of findings

Our results are consistent with previous studies that show that there might be an effect of accreditation on quality of care processes and outcomes, although effects are not always consistent[6,7,10]. Furthermore, we mainly found improvements over time on those measures that had a lower score at baseline, which is consistent with findings from a systematic review on quality improvement strategies[26]. Several scores, especially regarding diabetes, were already quite high, which might be a reflection of the general amount of attention towards diabetes care improvement, but it can also be related to the relatively large numbers of practices that provide vocational training in our study population, which is associated with better quality of care[27]. In the Netherlands, approximately 30% of all practices have at least one family physician that provides training. In our population the percentage of practices with training was higher in both cohorts: 61% (42 of 69 practices) and 70% (48 of 69 practices) in the first and second cohort respectively.

In the Dutch Accreditation program, the practice is accredited if it is actively involved in quality improvement through use of improvement plans and can achieve improvements over time. An advantage of this approach is that the program is more attractive for health professionals; they take ownership of the improvement plans that are tailored to the individual practices. Like other accreditation programs, it uses audit and feedback as a central mechanism, embedded in a continuous quality improvement process that aims to enhance a culture that fosters quality and safety of healthcare[4,28]. The fact that we did not find stronger evidence for an added value of the Accreditation program may be partially explained by other developments in the primary care field during our research period. Starting in 2007, bundled payment was implemented broadly for diabetes and to a more limited extent for COPD; implementation of bundled payment for CVD just started at the end of our research period (2011). Although a large part of the practices may currently participate in bundled payment schemes[15], during our study period, most schemes were only just set up. Bundled payment offers audit and feedback and funds for support and education, but offers no other interventions to achieve quality improvement. The Dutch Accreditation program is one example of an intervention that can complement further implementation through the development of improvement plans and continuous support. There are clear similarities between participation in bundled payment and the accreditation process as an audit and feedback system. Both require data collection and offer feedback; measures used on diabetes, COPD and CVD are almost identical. In fact, practices that participate in both initiatives often collect data only once and use the data for both purposes. This might explain why little added value of the Accreditation program was detected for diabetes and COPD, two conditions for which many practices also participate in a bundled payment initiative. It could also account for the effect of accreditation on CVD, since implementation of bundled payment for this condition only just started during our research period.

Strengths and limitations

Our study design implied a risk of confounding, because of the absence of random allocation. The second cohort was selected out of a possible 802 practices of all 4,917 Dutch practices (16.3%)[29]. At the end of our study period, more than half of all Dutch family physicians (over 4,600) worked in a practice that took part in the Accreditation program and more than 40% of all practices were accredited. It is likely that a substantial part of these practices also participated in bundled payment for diabetes and COPD; the lack of exact information on this limited this study. In our models, we accounted for practice type, size and location. However, there may have been an underrepresentation of practices with a less than average interest in quality improvement, especially in the first cohort, since participation in the program was voluntary. Both the argument of confounding and of the representation of the study

population reflect on the possibility that the effects could be more present in the whole population of general practices.
The increase in diabetes and COPD prevalence between 2006 and 2011 may indicate that practices have improved their registration. On the other hand it can also be a result of efforts to screen the population for undiagnosed patients. Process indicator scores might decrease when registration is better, since there is an association between performing a measurement and registering it. This can cause a bias, which may differ when registration improves. Outcome scores might increase after active screening of the population, due to the inclusion of patients with a mild form of diabetes or COPD.

Conclusions

General practices improved the quality of provided healthcare for patients with the targeted chronic conditions. Nevertheless, only a few improvements could be attributed to the Accreditation program, which may be caused by the fact that other programs address quality of chronic care at the same time. Expectations of the effects of accreditation were high among participants and stakeholders, but the results of this evaluation do not support these fully. Continuous monitoring may have been beneficial for practices in order to maintain a high level of quality in some areas and improve further in others. After the study was completed, adaptations to the program were made, which reduced the burden of work and may improve the effectiveness. For instance, these could stimulate practices to not focus solely on the measured items[30].

REFERENCES

1. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the Chronic Care Model in the new millennium. Health Aff 2009;28(1):75-85.
2. Shaw C. The external assessment of health services. World Hosp Health Serv 2004;40(1): 24-7, 50, 51.
3. Jaafaripooyan E, Agrizzi D, Akbari-Haghighi F. Healthcare accreditation systems: further perspectives on performance measures. Int J Qual Health Care 2011;23(6):645-56.
4. Alkhenizan A, Shaw C. Impact of accreditation on the quality of healthcare services: a systematic review of the literature. Ann Saudi Med 2011;31(4):407-16.
5. Shaw C, Groene O, Mora N, Sunol R. Accreditation and ISO certification: do they explain differences in quality management in European hospitals? Int J Qual Health Care 2010;22(6):445-51.
6. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, et al. Audit and feedback: effects on professional practice and health-care outcomes. Cochrane Database Syst Rev 2012;6:CD000259.
7. Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review. Int J Qual Health Care 2008;20(3):172-83.
8. Greenfield D, Braithwaite J. Developing the evidence base for accreditation of healthcare organisations: a call for transparency and innovation. Qual Saf Health Care 2009;18(3): 162-3.
9. Hinchcliff R, Greenfield D, Moldovan M, Westbrook JI, Pawsey M, et al. Narrative synthesis of health service accreditation literature. BMJ Qual Saf 2012;21(12):979-91.
10. O'Beirne M, Zwicker K, Sterling PD, Lait J, Lee Robertson H, et al. The status of accreditation in primary care. Qual Prim Care 2013;21(1):23-31.
11. Szecsenyi J, Campbell S, Broge B, Laux G, Willms S, et al. Effectiveness of a quality-improvement program in improving management of primary care practices. CMAJ 2011; 183(18):E1326-33.
12. in 't Veld C. NHG-Praktijkaccreditatie: een keurmerk voor ambitie. Bijblijven 2010;26(9):7-12.
13. website NHG-Praktijkaccreditering. http://www.praktijkaccreditering.nl.
14. de Wildt JE, Leusink GL. Nulmeting zorg-groepen: een beschrijvend onderzoek van de karakteristieken. Utrecht: LVG; 2008.
15. Landelijke Organisatie voor Ketenzorg. Transparante ketenzorg: diabetes mellitus, COPD en VRM. Rapportage zorggroepen over 2012. Utrecht: LHV/VHN/LVG; 2013.
16. Porter ME, Pabo EA, Lee TH. Redesigning primary care: a strategic vision to improve value by organizing around patients' needs. Health Aff 2013;32(3):516-25.
17. Hussey PS, Mulcahy AW, Schnyer C, Schneider EC. Closing the quality gap: revisiting the state of the science (vol. 1: bundled payment: effects on health care spending and quality). Evid Rep Technol Assess 2012;208.1:1-155.
18. de Bakker DH, Struijs JN, Baan CB, Raams J, de Wildt JE, et al. Early results from adoption of bundled payment for diabetes care in the Netherlands show improvement in care coordination. Health Aff 2012;31(2):426-33.
19. Struijs JN, Baan CA. Integrating care through bundled payments – lessons from The Netherlands. N Engl J Med 2011;364(11):990-1.
20. Struijs JN, van Til JT, Baan CA. Experimenting with a bundled payment system for diabetes care in the Netherlands: the first tangible effects. Bilthoven: RIVM; 2010.
21. Scott A, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database Syst Rev 2011;9:CD008451.
22. van den Hombergh P, Grol R, van den Hoogen HJ, van den Bosch WJ. Assessment of management in general practice: validation of a practice visit method. Br J Gen Pract 1998; 48(436):1743-50.
23. Grol R, Wensing M, Mainz J, Jung HP, Ferreira P, et al. Patients in Europe evaluate general practice care: an international comparison. Br J Gen Pract 2000;50(460):882-7.
24. Braspenning J, Dijkstra R, Tacken M, Bouma M, Witmer H. Visitatie Instrument Accreditering (VIA®). Nijmegen/Utrecht: Kwaliteit van Zorg (WOK); UMC St Radboud; Nederlands Huis-

artsen Genootschap (NHG); NHG Praktijk Accreditering BV (NPA); 2007.
25. IBM SPSS statistics for Windows, version 20.0. Armonk: IBM Corp; 2011.
26. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet 2012;379(9833):2252-61.
27. van den Hombergh P, Schalk-Soekar S, Kramer A, Bottema B, Campbell S, et al. Are family practice trainers and their host practices any better? Comparing practice trainers and non-trainers and their practices. BMC Fam Pract 2013;14:23.
28. Greenfield D, Pawsey M, Braithwaite J. What motivates professionals to engage in the accreditation of healthcare organizations? Int J Qual Health Care 2011;23(1):8-14.
29. van Hassel DTP, Kenens RJ. Cijfers uit de registratie van huisartsen – peiling 2012. Utrecht: NIVEL; 2013.
30. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ 2011;342:d3590.

5

Opinions of family physicians regarding the Dutch Accreditation program

Arna van Doorn - Klomberg
Kirsten Kirschner
Margriet Bouma
Kees in 't Veld
Michel Wensing
Jozé Braspenning

Published as: 'Ervaringen met de NHG - Praktijkaccreditering®' in Huisarts en Wetenschap 2011; 54(7): 360-5.

ABSTRACT

Background In this study, we describe the opinions of family physicians regarding the Dutch Accreditation program of the Dutch College of General Practitioners.

Methods Using a mixed methods design, data were collected by means of questionnaires and interviews from 61 and 29 family physicians respectively, as part of the Dutch Accreditation Program in 2006 and 2007. The physicians also took part in a pay-for-performance study, where they received a progressive bonus, based on their performance.

Results The family physicians chose to participate in the Dutch Accreditation program mainly because they wanted to improve the quality of care. While they were positive about the insight they gained into their practice management and clinical care, they found that the collection of data on clinical care was too time-consuming and that they needed more support in preparing a practice improvement plan. They also thought that the assessment instruments should put less emphasis on details and more on the underlying processes.

Conclusions Family physicians are willing to take part in an accreditation program, but the process should not be too time-consuming and the instruments used should lend themselves to improving quality now and in the future. It is important to find a balance between measurability and perceived quality. Some physicians considered the drawbacks of the process to outweigh its advantages.

BACKGROUND

Periodic screening of practice performance and accountability are part of a professional approach to healthcare. Furthermore, in the (re)organization of the Dutch healthcare system, the importance of transparency is emphasized. This has created an increased need to gain insight into medical care and practice management, as well as which aspects should be improved. Accreditation of general practices offers a method by which to obtain this information[1,2]. In accreditation, dynamic targets are used to assess whether quality of care is improved in a correct manner[3]. This continuous, systematic improvement process fits well within the setting of general practice care. In the Netherlands, the Dutch Accreditation program has been developed, in which practices can participate since 2005. The Visitation Instrument Accreditation (VIA) that is used in the program (see Box 1) covers several domains: practice management, clinical care and patient experiences[4]. The indicators that are used in this instrument are sufficiently valid and reliable[5-7].

Box 1. Dutch Accreditation program and Visitation Instrument Accreditation (VIA)[8-10]

The Dutch Accreditation program is a process with a three-year cycle. The practice collects data using the Visitation Instrument Accreditation (VIA) and receives feedback through a feedback report with benchmarks. Based on this feedback, the practice develops SMART improvement plans[11] and implements these. In order to get accredited, the practice must meet certain basic requirements, for example regarding hygiene. In addition, the accreditor assesses the quality of the improvement plans and verifies whether the practice is actively improving the quality of care. The VIA consists of performance scores in three different domains: clinical care, practice management and patient experiences. The data are collected with validated questionnaires filled in by patients, family physicians (FPs), practice assistants, nurse practitioners and the consultant who visits the practice. Clinical care data are mainly extracted from the medical records. Most clinical care performance measures concern diabetes, chronic obstructive pulmonary disease (COPD), asthma and cardiovascular diseases. If the practice is not able to extract the data automatically, a randomly selected sample of forty patients per condition can be taken instead. Validity and reliability have been assessed during the development of the instrument.[5-7]

In 2006 and 2007, 65 practices participated in a pay-for-performance program (see Box 2), which included participation in the Dutch Accreditation program[12]. In the pay-for-performance program, quality of care was compared among the participants. The practices received a bonus for participation[13,14]. As part of the evaluation of the pay-for-performance program, the Dutch Accreditation program was also evaluated. Insight in the motivation and experiences of the participants is crucial to maximize the effects of a program like the Dutch Accreditation program[15]. Greenfield et al. found that participation of professionals in an accreditation scheme within a hospital setting in Australia contributed to a positive change in the team climate, which led to an increased focus on quality and safety[15]. To our knowledge, such research has not been done yet in the Dutch primary care setting. With this study we aim

Box 2. The pay-for-performance program[12-14]

Family physicians (FPs) in the south of the Netherlands could voluntarily participate in a study on the effects of awarding a financial bonus based on performance. A bonus system was used in which practices received points for their scores on the performance measures. For each measure, points could vary between 0 and 6; the division into groups was based on the percentile scores of all practices. The practice received a financial bonus per patient. An average practice received a fee of approximately 7,500 euros. The fee per practice varied from 0 to 15,000 euros. Costs for participation for an average practice were approximately 6,200 euros. Some of the practices received a fee that was lower than the participation costs. At the time of conducting the interviews, not every FP knew how high the bonus would be. However, all family physicians were familiar with the bonus system used. Participation in the Dutch Accreditation program was a condition for participation in the pay-for-performance program. Some practices already participated in the Dutch Accreditation program before they started the pay-for-performance program. Several practices started participation in the Dutch Accreditation program to be able to participate in the pay-for-performance program. However, for most practices no information is available on whether they already participated in the Dutch Accreditation program at the start of the pay-for-performance program.

to gain insight into the experiences of family physicians who participate in the Dutch Accreditation program.

METHODS

Study design

We used a mixed methods design[16,17] to include a broad range of aspects related to the experiences of the family physician (FP). We looked at whether their experiences met their expectations through a quantitative questionnaire which can be found in the Appendix. Results from the questionnaires were refined and elaborated on through a qualitative interview study.

Questionnaires

As part of the pay-for-performance program, we distributed an evaluative questionnaire among 65 practices in the southern part of the Netherlands. All practices collected data in 2006 and 2007 for the Dutch Accreditation program and participated in the pay-for-performance program[12]. An overview of the practice characteristics can be found in Table 1. The questionnaires consisted of open-ended and closed-ended questions; part of the questions used a 5-point Likert scale.

Table 1. Practice characteristics of the two study populations (questionnaires/interviews) compared to all Dutch practices

Practice characteristics	Study population questionnaires (n = 65)	Study population interviews (n = 29)	All practices (n = 4321)
Practice type			
Single-handed	29.2%	17.9%	43.5%*
Two FPs	30.8%	39.3%	31.8%
Group practice or health center	40.0%	42.8%	24.7%*
Practice location◊			
Urban	27.9%	31.0%	46.6%*
Partly urban	60.6%	62.1%	41.1%*
Rural	11.5%	6.9%	12.3%*
Practice size			
Average no. of patients per practice	4596	5076	4283‡
Workload			
Average no. of FTE FPs per 1000 patients	0.41	0.42	0.42^

** NIVEL, 1-1-2007[18]. ◊ A region was defined as urban when the number of addresses per km² exceeded 1,500; partly urban regions had 500 - 1.500 addresses per km²; rural regions had less than 500 addresses per km². ‡ Second National Study[19]. ^ VIA 2006/2007[20].*

Interviews

Three trained interviewers conducted semi-structured interviews with part of the FPs that filled in the questionnaire. We approached all physicians; of the 35 FPs who were willing to participate, we approached the 30 FPs who responded first. In this qualitative study, we used the framework approach[21,22]. For the design of the interviews and questionnaires we constructed a framework of a number of themes with subthemes in advance (Table 2). No new questions were added during the conduction of the interviews.

Analysis

The completed questionnaires were analyzed using descriptive statistics. In order to understand the impact of the financial bonus on the motivation and experiences of the participants, we looked at the differences in the scores of two groups, classified according to size of their bonus per patient. The data from the interviews were analyzed using a thematic analysis[16]. Not all the subthemes that were established in advance were equally important to the participants; new themes emerged during the analysis. We inductively defined these themes, see the bold (sub)themes in Table 2. The information from the interviews was placed within the framework. Two researchers independently assessed the interviews in order to guarantee the reliability of the analysis. Discrepancies were discussed and consensus was reached in all cases.

Table 2. Overview of themes and subthemes of the interviews and questionnaires; bold (sub)themes have been included inductively during data analysis

Theme	Questionnaires	Interviews
Motivation for participation	x	x
Data collection		
Data from medical records		x
Data from questionnaires		x
Data from observation		x
Opinion on data collection in general	x	x
Time investment	x	x
Contents of the instrument (VIA)		x
Improvement plans		
Opinion on improvement plans in general	x	x
Time investment	x	x
Support		
Applying for participation		x
Practice consultant		x
VIA-consultant		x
Accreditor		x
Assistance with data collection		x
Support in drafting improvement plans		x
Benefits for the practice		
Opinion on participation in general	x	x
Data collection		x
Changes in the team		x
Financial aspects		x

Table 3. Motivation to participate in the Dutch Accreditation program

Reason for participation	No. of FPs (out of 61 FPs)
Improve patient care	50
Gain insight into clinical care	47
Gain insight into practice management	42
Provide structure to implement changes in the practice	33
Gain insight into patient experiences	30
Show others what we do, because we are proud of that (transparency)	27
Participation was a condition for/to… (with explanation)	21
Improve market and bargaining power of the practice	16
Other reason, namely… (with explanation of the practice)	8

RESULTS

Study population

Out of the 65 practices we approached, 61 family physicians (FPs) from as many practices completed the questionnaire. Of the 30 FPs that were approached for an interview, 29 took part. For one FP it was not possible to make an appointment for the interview during the study period.

Motivation for participation

In the questionnaires, the FPs gave various reasons for participation in the Dutch Accreditation program (see Table 3). The arguments that were mentioned most frequently included improving patient care and gaining insight into their clinical care and management. Out of 61 practices, 21 indicated that they participated in the Dutch Accreditation program because it was a requirement for participation in the pay-for-performance program. The amount of the bonus had no effect on the reasons for participation.

Time investment

Results from the questionnaire showed that the amount of time that physicians had to invest in order to collect data was much greater for chronic conditions than for other topics (Table 4). Even if the FPs were able to extract data from their electronic medical records (EMR), practices often needed more than 4 hours to collect data. At the time of this study, the practices did not gain time by automated data collection. In the interviews, FPs reported that the collection of data on clinical care took a lot of time, especially since it was not possible to easily retrieve the data from medical records. Several FPs indicated that EMR software suppliers should make adjustments for this purpose. Sometimes the requested information was available via a secondary source, but it either turned out to be laborious to provide this information, or the information was not accessible to the FPs. FPs also indicated that they felt that data collection was not part of their job responsibilities. For example, one of the FPs said: 'The quality of care, the actual time spent with patients, is surely the most important. Now you have to fill in paperwork. Things that should not really be done by a physician – data collection, getting information – Instead of seeing a patient on Monday, I was busy with paperwork'.

Table 4. Time investment of the data collection per subject and per practice

Subject*	>4 hours	2-4 hours	<2 hours
Diabetes, COPD, asthma, or cardiovascular diseases	58%	32%	10%
Medication or prevention	17%	28%	54%

** Because the percentages were rounded off, they do not always add up to exactly 100%.*

Contents of the VIA instrument

In general, results from the interviews show that the FPs were positive about the contents of the practice management domain, particularly the inspection of supplies and instruments such as the content of the doctor's emergency bag. However, several of the interview participants felt that they were judged on things that they found less important. They were afraid that with the scoring of checklists, the underlying concept, the quality of care, would disappear into the background. Some physicians thought that several of the indicators were not directly related to the quality of care. One of the physicians mentioned an example: 'The date of the ampoules. Of course we have that in order. You just know that they will check that. You arrange that a day in advance. Those are all little things that actually do not matter in my opinion. What matters is a broader picture... how do you guarantee certain things? You can think about improvement now, but how do you guarantee that for the future? After all, you have to ensure that you maintain your quality level. And not just make an effort for the sake of those improvement plans in one year and then let it slip'. Furthermore, FPs felt that some measurements were too strict. For example, practices scored negatively if they followed a successful work process but did not have this recorded in a written protocol. One of the participants doubted whether the information supplied by the patients was accurate. However, results from the interviews showed that physicians felt it was important to include measurements from the patient's perspective. FPs were satisfied with the diversity of the VIA instrument, both in terms of the subjects, as well as the sources for data collection. Nevertheless, they thought that extra subjects could be added to the instrument, for example innovation and cooperation with other practices within the care group.

Improvement plans

In the questionnaire, the majority of FPs (88%) indicated that it was difficult to write improvement plans. 44% of the FPs thought the stated requirements were clear. On average they spent 18.5 hours on writing the improvement plans. More than half of the FPs indicated in the questionnaire that the improvement plans contributed to the quality of care. The interviews revealed that the entire process of writing and approving the improvement plans took a lot of time, which stalled the process. FPs would have preferred to spend less time on this, but also indicated that it was beneficial for them to make the plans themselves. They experienced the process as too bureaucratic and they reported that relatively little attention was paid to the content. For example, one FP said: 'At first I found it rather difficult that there was a strong focus on the (required) format. I think the contents are fine. I like to refine things further, but in this case I found that sometimes there was a bit too much talk just for the sake of it. It wasn't nonsense, but I did often think that the format sometimes seemed more important than our ultimate goal... We're very pragmatic'. A number of FPs also indicated that

a strict method provides the guidance needed to achieve feasible improvement plans. However, they did want more support. The only support they could obtain had to be funded by themselves. They thought that was not sufficient.

Implementation support

In the interviews, FPs reported that the accreditation process should be more streamlined and that the communication needed improvement. Especially at the start of the accreditation the process did not go smoothly. Results from the interviews showed that there were differences in the approach of the accreditors, which also applied to the consultants. Furthermore, we found that FPs sometimes felt the need to compare themselves with other practices in the area. The image that they had of these practices did not always correspond with the observation of the accreditor. For example, one FP responded: 'When I see how easily some practices have received their accreditation and I compare our practice with theirs, I think that it is not equivalent. We might both be accredited, but the difference in organization is apparently not expressed in the accreditation. If you have a lot to improve, then you also have improvement plans quickly and you can also improve three things quickly. I think that there should be (more) minimum requirements that should be met'. Furthermore, different insurers had different compensation systems, which the FPs judged negatively.

Benefits for the practice

The questionnaire showed that the majority of FPs thought that the time investment to take part in the Dutch Accreditation program was worthwhile and for almost all FPs the goal of the program was clear. About half of the FPs would recommend a colleague to participate, 20% would advise against it. The remaining approximate 30% scored neutral on this question. The responses with regards to whether their experiences were consistent with their expectations of the Dutch Accreditation program were also very diverse. The relative size of the bonus did not influence their opinions. The group with the relatively large bonus seemed to score slightly more positive on the questions 'participation is worth the time investment' and 'would you recommend a colleague to participate?', in comparison with the group with the relatively small bonus. The differences were not significant; however, small numbers of physicians were included per group. In the interviews, FPs indicated that during the data collection they gained more insight in their own performance. Part of the issues that came up could be addressed directly. Several FPs reported that they performed worse than they had anticipated. One physician responded: 'In principle I am excited about the accreditation, because it provides insight. You always think you're already doing very well and then you find out that some things really can be improved upon... And there are issues that come up that can easily be improved'. The FPs were provided with a broad view of their practice

management on paper. This made it easier to start working on concrete improvements. According to the FPs, participation also had a positive impact on the cooperation and the atmosphere within the team. Some FPs indicated that the enormous amount of work did not outweigh the benefits of improvement and insight. Several FPs thought that the accreditation could be particularly valuable for FPs who have much to improve. According to them, FPs who are already at the forefront in their field have less to improve. Physicians also evaluated the recognition of the Dutch College of General Practitioners (which consisted of promotion material that included a certificate for each FP in the practice, leaflets and stickers) as very poor in comparison with the time and money investment of the practice. However, several practices mentioned that they were now more active in quality improvement due to the Accreditation program.

DISCUSSION

The results show that actively working on quality improvement is the main motivation for family physicians to participate in an accreditation program such as the Dutch Accreditation program. Benefits for the physicians include more insight in their own practice management and the possibility to work on quality improvement with improvement plans that are drawn up by the practice staff itself. FPs noticed that some things already improved during the preparation for the visitation. According to the FPs, the main points to improve upon included the amount of time it took to collect data, particularly on clinical care, and the preparation of improvement plans. They would have also liked to receive proper and free support. FPs prefer to pay less attention to details and more to the underlying processes, so that continuous quality improvement is indeed a continuous process. For part of the FPs the disadvantages of the process of the Dutch Accreditation program outweigh the benefits of participation.

Strengths and limitations

The study population consisted of FPs who also took part in the pay-for-performance program. Out of 61 practices, 21 practices indicated to participate in the Dutch Accreditation program because that was required. The cash bonus possibly gave FPs a greater sense of recognition, which may have resulted in a more positive judgment. On the other hand, they may have been more critical about the contents of the indicators and the method of measurement (self-assessment) because of the increased importance of the measurements. In the selection of the FPs for the interviews, we did take into account the size of the financial bonus (we made a proportional distribution). We did not find any significant differences between these groups, probably because of the small number of participants. The evaluative questionnaires and interviews mainly concerned questions regarding the pay-for-performance program. Also, the main focus of the interviews was on clinical care, which was

the most relevant part according to the FPs. Therefore, the other components possibly have not been given sufficient attention. There are several limitations to the method that we chose for this study. During the period that the interviews took place, we did not add any new themes (inductively derived), therefore several themes might not have been sufficiently addressed in the interviews. We have conducted the interviews among a limited number of FPs; however, during the period of interviewing saturation was reached[23].

Implications

The experiences of FPs are associated with the motivation and commitment to participate, and thus with the effects that can be achieved through accreditation. First, good communication on the development of indicators and their purpose is vital. Some FPs indicated that the instrument came across as a checklist of minor things, that was not actually about 'quality'. A better understanding of the development of an indicator set and its background can help participants to better understand what is being measured, which in turn might increase their motivation. Regarding several aspects that are related to 'quality', such as communication with patients, no clear criteria are available. This makes it difficult to create measurable indicators for these aspects. There is an area of tension between measurability and what is actually perceived as 'quality'. Performance measures should be seen as a tool for quality improvement. Second, the experiences with data collection are important. Although data collection is laborious and FPs feel that this is not part of their job responsibilities, it can, however, contribute to quality improvement when FPs collect (part of) the data themselves. FPs will be less likely to dismiss the results as inaccurate and irrelevant. Also, during the data collection FPs often gain insight in their own actions: it regularly turns out that reality is less favorable than expected[24]. A third point concerns the writing of improvement plans. Although a SMART approach is used within the Dutch Accreditation program to formulate goals for improvement plans[25], this approach did not always suit the working method of FPs well, which sometimes led to frustration. To keep all FPs motivated during an accreditation process, it is important not to focus too much on the methodology, but to aim for a substantive and pragmatic approach, without losing the underlying method of concrete and comprehensive formulation. Fourth, physicians indicated that the assessment by accreditors was not uniform. However, they also suggested that accreditors should assess the individual situation, so that alternative solutions that do not comply with the guidelines can still get approved. Here is an area of tension: if strict assessment rules are set for consultants and accreditors, their assessments will be less diverse, but this also leads to less flexibility in the assessment of practices. Within the NPA, the company that manages the Dutch Accreditation program, several steps have been taken to standardize their methods since conducting the interviews. In the certification scheme, the emphasis remains on

systematic quality improvement based on performance measurement. The practice itself is responsible to choose a good solution, tailored to the practice, for existing risks of safety and quality.

Conclusions

Family physicians who also participated in the pay-for-performance program were somewhat positive about the Dutch Accreditation program. They deem further development of the Accreditation program necessary; which is in line with the continuous development of the program within the NPA[11]. We can draw several broader conclusions based on our results. Physicians are willing to evaluate their own practice management and clinical care, using an accreditation process. It is important that the data collection is not too time consuming – the large time investment formed the main barrier in the Dutch Accreditation program. It is also important that the instrument that is used encourages real quality improvement, which can be ensured in the future. This involves areas of tension between measurability and perceived 'quality', and between flexibility and consistency in assessment.

REFERENCES

1. Sutherland K, Leatherman S. Regulation and quality improvement: a review of the evidence. London: The Health Foundation; 2006. www.health.org.uk/QQUIP.
2. Grol R, Wensing M (redactie). Implementatie: effectieve verbetering van de patiëntenzorg. Maarssen: Elsevier gezondheidszorg; 2006 (3e druk).
3. Pauwels J. Kwaliteit kenbaar maken in de zorg. 2010. www.zorgnetvlaanderen.be/Documents/2010%20Accreditering.pdf.
4. Braspenning J, Dijkstra R, Tacken M, Bouma M, Witmer H. Visitatie Instrument Accreditering (VIA®). Nijmegen/Utrecht: Kwaliteit van Zorg (WOK); UMC St Radboud; Nederlands Huisartsen Genootschap (NHG); NHG Praktijk Accreditering BV (NPA); 2007.
5. van den Hombergh P, Grol R, van den Hoogen HJ, van den Bosch WJ. Assessment of management in general practice: validation of a practice visit method. Br J Gen Pract 1998; 48(436):1743-50.
6. van Doorn A, Kirschner K, Bouma M, Burgers J, Braspenning J, et al. Evaluatie van het onderdeel medisch handelen van de accreditering: vier klinimetrische criteria. Huisarts Wet 2010; 53(3):141-6.
7. van den Hombergh P, Grol R, Van den Hoogen HJ, Van den Bosch WJ. Practice visits as a tool in quality improvement: acceptance and feasibility. Qual Health Care 1999;8(3):167-71.
8. Witmer H, Bouma M, Braspenning J, in 't Veld C, Grol R. NHG Praktijkaccreditering®. Een nieuwe stap in de ontwikkeling van het kwaliteitsbeleid van huisartsenpraktijken. Utrecht/Nijmegen: NHG/WOK, UMC St. Radboud; 2005.
9. van den Hombergh P. Practice visits. Assessing and improving management in general practice. Nijmegen: Kwaliteit van Zorg (WOK), UMC St Radboud; 1998.
10. Wensing M. Patients evaluate general practice care. Nijmegen: UMC St. Radboud; 1997.
11. NPA beleidsnotitie; 2008. http://npa.artsennet.nl/professionals/huisartsen/belangrijke-documenten.htm.
12. Kirschner K, Braspenning J, Gootzen T, Van Everdingen C, Batenburg J, et al. Pay-for-performance in de huisartsenpraktijk: een experiment in Zuid-Nederland. Nijmegen: IQ healthcare, Radboudumc; 2009.
13. Braspenning J, Kirschner K, Batenburg J, van de Rijt D, Grol R. Loon naar werken loont. Med Contact 2008;63:1042-5.
14. Kirschner K, Braspenning J, Batenburg J, van de Rijt D, Muijers P, et al. Value for money: een model voor honoreren van kwaliteit in de huisartsenpraktijk. Nijmegen: Kwaliteit van Zorg, UMC St Radboud; 2008.
15. Greenfield D, Pawsey M, Braithwaite J. What motivates professionals to engage in the accreditation of healthcare organizations? Int J Qual Health Care 2011;23(1):8-14.
16. Pope C, Mays N. Qualitative research in health care. Oxford: Blackwell Publishing; 2006 (3rd ed.).
17. Pope C, Mays N. Critical reflections on the rise of qualitative research. BMJ 2009;339:b3425.
18. website NIVEL-beroepenorganisaties; 2007. www.nivel.nl.
19. Schellevis EG, Westert GP, de Bakker DH, Groenewegen PP. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Utrecht/Bilthoven: NIVEL/RIVM; 2004.
20. Visitatie Instrument Praktijkaccreditering, gegevens NPA, 2006/2007.
21. Pope C, Ziebland S, Mays N. Analysing qualitative data. BMJ 2000;320(7227):114-6.
22. Richie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess RG (editors). Analyzing qualitative data. London: Routledge; 1994: 173-94.
23. Boeije H. Analyseren in kwalitatief onderzoek: denken en doen. Amsterdam: Boom; 2005.
24. Braspenning J, Hermens R, Wollersheim H, Grol R. Meten van (veranderingen in) de zorg: de rol van indicatoren. In: Grol R, Wensing M (redactie). Implementatie: effectieve verbetering van de patiëntenzorg. Maarssen: Elsevier gezondheidszorg; 2006 (3e druk): 153-79.
25. Doran GT. There's a S.M.A.R.T. way to write management's goals and objectives. Management Rev 1981;70(11):35-6.

APPENDIX

This appendix contains a translated version of the written questionnaire that was used in this study to evaluate family physician experiences. The instructions can be found in Box 1.

Box 1. Instructions questionnaire

You have recently participated in the Dutch Accreditation program as part of the pay-for-performance program for family physicians. Through this questionnaire, we would like to gain more insight into your personal experiences with the Dutch Accreditation program.

As you may know, the Dutch Accreditation program aims to evaluate:

- ✓ *medical care*
- ✓ *practice organization, operational management and finance*
- ✓ *patient experiences*

Based on the data collected in these three domains, your practice has received feedback and improvement plans have been evaluated by an accreditor. Accreditation is granted in the first year when:

- ✓ *the accreditor reports a positive advice*
- ✓ *the list of basic information regarding the practice is accurate and up-to-date.*

The Accreditation program uses a three-year cycle. After being accredited in the first year, the evaluation of quality improvements provides the basis on which a practice can be accredited in the second and third year.

This questionnaire is intended for the quality coordinator of the practice. We would like to ask you several questions regarding the Dutch Accreditation program. The main topics of this questionnaire include motives for participation, the registration process, data collection, feedback, improvement plans and support from the Dutch College of General Practitioners (NHG/NPA).

1. Personal background

1) what is your gender: male/female
2) what is your age: _____ years
3) what is your education: _____
4) what is your current job function: _____

2. Motives for participation

Which motive(s) did you have to participate in the Dutch Accreditation program? (multiple answers possible)

- □ improve patient care
- □ gain insight into clinical care
- □ gain insight into practice management
- □ gain insight into patient experiences
- □ improve market and bargaining power of the practice
- □ provide structure to implement changes in the practice
- □ show others what we do, because we are proud of that (transparency)
- □ participation was a condition for/to _____
- □ other reason, namely _____

3. Registration with the Dutch College of General Practitioners

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1) the registration process went smoothly	□	□	□	□	□
2) after reading the information material, the procedure of the program was clear	□	□	□	□	□

3) do you have suggestions for improvement? _____

4. Data collection

You have collected data regarding three different domains: medical care, practice organization and patient experiences. There were different methods of data collection within each domain, therefore the questions are formulated for each of the domains separately.

4.1 Medical care

1) how many family physicians work in the practice? _____ physicians
2) possibly there are several separate patient populations within the practice. Could you clarify which family physicians cover the same patient population?
   patient population 1: _____
   patient population 2: _____
   patient population 3: _____
   patient population 4: _____
   patient population 5: _____

3) how many separate data collections of medical care have taken place within the practice? *(in theory, the practice should have collected data for each of the patient populations separately in case of more than 1 patient population)* _____ data collections

4) has the data been collected for a randomly selected sample of all patients, or could data be collected on the entire patient population (select both in case different methods were used for different populations)

	Selected sample	Complete population
Diabetes	□	□
COPD	□	□
Asthma	□	□
Cardiovascular diseases	□	□
Prevention (influenza, cervical screening)	□	□
Medication prescription	□	□

5) which data source was used to collect the data (select both in case different methods were used for different populations)

	Practice medical records	Other source
Diabetes	□	□
COPD	□	□
Asthma	□	□
Cardiovascular diseases	□	□
Prevention (influenza, cervical screening)	□	□
Medication prescription	□	□

6) how much time was spent in total on data collection?

	0-1 hour	1-2 hours	2-3 hours	3-4 hours	>4 hours
Diabetes	□	□	□	□	□
COPD	□	□	□	□	□
Asthma	□	□	□	□	□
Cardiovascular diseases	□	□	□	□	□
Prevention	□	□	□	□	□
Medication prescription	□	□	□	□	□

7) who was involved in the medical care data collection process?

	Family physician	Assistant	Nurse Practitioner	Other (e.g. pharmacist)
Diabetes	□	□	□	□
COPD	□	□	□	□
Asthma	□	□	□	□
Cardiovascular diseases	□	□	□	□
Prevention	□	□	□	□
Medication prescription	□	□	□	□

8) the data collection sheets that needed to be filled in by the practice were clear:

strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
□	□	□	□	□

9) were there any factors that made the data collection regarding medical care more difficult? yes/no
If so, which factors? _____

10) were there any factors that made the data collection regarding medical care easier? yes/no
If so, which factors? _____

11) do you have suggestions for improvement? _____

4.2 Practice organization

1) how much time was spent in total to fill in all questionnaires regarding practice organization?
_____ , _____ hours/minutes

2) who was involved in the practice organization data collection process?

Assistant	Nurse Practitioner	Family physician	VIA consultant
☐	☐	☐	☐

3) were there any factors that made the data collection regarding practice organization more difficult?
yes/no If so, which factors? _____

4) were there any factors that made the data collection regarding practice organization easier?
yes/no If so, which factors? _____

5) the data collection questionnaires that needed to be filled in by the practice were clear:

strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
☐	☐	☐	☐	☐

6) do you have suggestions for improvement? _____

4.3 Patient experiences

1) how much time was spent in total to hand out and retrieve questionnaires regarding patient opinions?
_____ , _____ hours/minutes

2) who was involved in the patient questionnaire data collection process?

Assistant	Nurse Practitioner	Family physician
☐	☐	☐

3) do you have suggestions for improvement? _____

5. Feedback

Based on the data you have collected, a feedback report was sent for all domains in the evaluation. Since these reports differ for each domain, the questions are again formulated for each of the domains separately.

5.1 Medical care

1) the feedback on medical care was appreciated by the practice:

strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
☐	☐	☐	☐	☐

2) how much time was spent in total to read the information in the feedback report on medical care?
_____ , _____ hours/minutes

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
3) the text in the feedback report was clear	□	□	□	□	□
4) the scores in the feedback report were clear	□	□	□	□	□
5) the information in the report was complete	□	□	□	□	□
6) the report provided added insight into your performance regarding medical care	□	□	□	□	□
7) you intend to make changes based on the results from the feedback report	□	□	□	□	□

5.2 Practice organization

1) the feedback on practice organization was appreciated by the practice

strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
□	□	□	□	□

2) how much time was spent in total to read the information in the feedback report on practice organization?
_____ , _____ hours/minutes

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
3) the scores in the feedback report were clear	□	□	□	□	□
4) the information in the report was complete	□	□	□	□	□
5) the report provided added insight into your performance regarding practice organization	□	□	□	□	□
6) you intend to make changes based on the results from the feedback report	□	□	□	□	□

5.3 Patient experiences

1) the feedback on patient experiences was appreciated by the practice

strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
□	□	□	□	□

2) how much time was spent in total to read the information in the feedback report on patient experiences?
_____ , _____ hours/minutes

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
3) the scores in the feedback report were clear	□	□	□	□	□
4) the information in the report was complete	□	□	□	□	□
5) the report provided added insight into patient experiences	□	□	□	□	□
6) you feel that feedback on patient experiences is valuable for quality improvement	□	□	□	□	□

6. *Improvement plans*

Besides the data collection in three different domains, practices had to develop improvement plans as part of the Dutch Accreditation program.

□ the practice has not developed improvement plans → you can continue with section 7.

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1) the development of improvement plans was difficult	□	□	□	□	□
2) the requirements that were set for improvement plans were clear	□	□	□	□	□

3) how much time was spent in total on the development of improvement plans?
_____ , _____ hours/minutes

4) were there any factors that made the development of improvement plans more difficult? yes/no
If so, which factors? _____

5) were there any factors that made the development of improvement plans easier? yes/no
If so, which factors? _____

6) to which extent did the improvement plans contribute to quality improvement?

very much	much	neither much nor little	little	very little
□	□	□	□	□

7. *Support*

The Dutch College of General Practitioners offers support to the practices in several ways. First of all, the practice employs a practice consultant who informs the practice on the process of accreditation and helps with the development of the improvement plans. Second, there is a VIA-consultant that discusses the feedback report with the practice and offers suggestions for topics for improvement plans. And third, the accreditor assesses the improvement plans and provides advice to the Dutch College of General Practitioners on whether or not a practice should receive accreditation.

→ *If you have not yet been visited by an accreditor, you can skip section 7.3.*

7.1 Practice consultant

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1) it was easy to employ a practice consultant	□	□	□	□	□
2) the practice consultant that helps with the development of improvement plans is competent	□	□	□	□	□
3) the practice consultant that helps with the implementation of improvement plans is competent	□	□	□	□	□

7.2 VIA-consultant

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1) the VIA-consultant was appointed quickly	□	□	□	□	□
2) the VIA-consultant was competent in explaining and discussing the feedback report	□	□	□	□	□
3) the VIA-consultant was competent in suggesting topics for improvement plans	□	□	□	□	□

7.3 Accreditor

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1) the accreditor was competent in judging the improvement plans	□	□	□	□	□
2) the accreditor that offered advice on whether or not a practice is accredited was competent	□	□	□	□	□

8. General questions

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
1) participation was worth the time investment	□	□	□	□	□
2) the goals of the program were clear	□	□	□	□	□
3) the experiences of participation were in accordance with your expectations	□	□	□	□	□
4) you would advise your colleagues to participate	□	□	□	□	□

Other remarks _____

6

Patient characteristics associated with measurement of routine diabetes care: an observational study

Arna van Doorn - Klomberg
Jozé Braspenning
Femke Atsma
Birgit Jansen
Margriet Bouma
René Wolters
Michel Wensing

Plos One 2015; 10(3): e0121845.

ABSTRACT

Background Non-modifiable patient characteristics, including age, gender, ethnicity as well as the occurrence of multi-morbidities, are associated with processes and outcomes of diabetes care. Information on these factors can be used in case-mix adjustment of performance measures. However, the practical relevance of such adjustment is not clear. The aim of this study was to assess the strength of associations between patient factors and diabetes care processes and outcomes.

Methods We performed an observational study based on routinely collected data of 12,498 diabetes patients in 59 Dutch primary care practices. Data were collected on patient age, gender, whether the patient lived in a deprived area, body mass index and the co-occurrence of cardiovascular disease, chronic obstructive pulmonary disease, depression or anxiety. Outcomes included 6 dichotomous measures (3 process and 3 outcome related) regarding glycosylated hemoglobin, systolic blood pressure and low density lipoprotein-cholesterol. We performed separate hierarchical logistic mixed model regression models for each of the outcome measures.

Results Each of the process measure models showed moderate effect sizes, with pooled areas under the curve that varied between 0.66 and 0.76. The frequency of diabetes related consultations as a measure of patient compliance to treatment showed the strongest association with all process measures (odds ratios between 5.6 and 14.5).The effect sizes of the outcome measure models were considerably smaller than the process measure models, with pooled areas under the curve varying from 0.57 to 0.61.

Conclusions Several non-modifiable patient factors could be associated with processes and outcomes of diabetes care. However, associations were small. These results suggest that case-mix correction or stratification in assessing diabetes care has limited practical relevance.

BACKGROUND

Many quality programs for diabetes care aim at improving clinical processes and outcomes[1]. An important part of the goals set in these programs is the prevention of cardiovascular risk[2,3]. Relevant intermediate outcomes include glycated hemoglobin (HbA1c), blood pressure and serum cholesterol values. These outcome measures are associated with a range of non-modifiable patient characteristics, including gender, age, ethnicity and other socio-demographic factors[4-7]. For example, women in diabetes care have higher blood pressure and serum cholesterol levels compared to men[7]. In order to compare performance measures between practices, adjustment is suggested for these factors, as they could be unequally distributed in the different populations assessed[8]. Zaslavsky et al.[9] showed that adjustment for socio-demographic factors had little impact on most performance measures, but larger impact on a few measures, particularly when the measure involved a relatively large percentage of patients from a minority ethnic group. However, case-mix adjustment is not consistently applied, possibly due to differential perspectives and uncertainties about when and how to adjust[10].

Besides socio-demographic factors, differences in the severity of the condition can be associated with diabetes control as well. The severity of diabetes mellitus can be defined as the co-occurrence of other conditions. In this article we focus on multi-morbidity, which can be defined as 'any co-occurrence of medical conditions within a person'[11]. The number of patients with multi-morbidity has risen enormously in the past decades[12]. A systematic review[13] showed that conditions that often occur in diabetes patients include cardiovascular diseases (CVD), depression and cancer. These conditions may be directly related to diabetes (e.g. CVD) or indirectly related (e.g. chronic obstructive pulmonary disease, COPD)[14]. Several studies have shown that the presence of different types of multi-morbid conditions (in several patient groups, including patients with diabetes) can lead to better cardiovascular risk control, both regarding processes and outcomes of care[15-17]. Multi-morbidity may lead to lowered patient compliance with treatment and advice[18,19]. Turner et al.[20] found that the relationship between multi-morbidity and healthcare outcomes may vary, depending on how closely related the conditions are. They suggest that related multi-morbid conditions may lead to improved care delivery, whereas unrelated multi-morbid conditions may have a negative impact on health outcomes[20]. Voorham et al.[21] specifically looked at the effects of multi-morbidity on medication treatment and found that diabetes-related multi-morbidity could induce treatment intensification for cardiovascular risk factors.

Information on multi-morbidity can be used in case-mix adjustment of performance measures. However, the strength of the association between multi-morbidity factors, other case-mix factors and measures of processes and outcomes of care is not clear. Therefore, the

aim of this study was to identify: (I) which patient factors and prevalent multi-morbid conditions are related to processes and outcomes of diabetes care, (II) how strong these associations are, and (III) the relative importance of the factors identified as predictors.

METHODS

Study design and study population

We performed an observational study based on routinely collected data from Dutch primary care practices. In the Netherlands, routine diabetes care is primarily provided in general practice; therefore, in our study we focus solely on primary care. All data were collected as part of a national representative network of practices (LINH), in which data were extracted from the electronic medical records. In total, a sample of 12,498 diabetes patients from 59 general practices was included in this study. All practices extracted their data in the year 2010. The extraction included information from all contacts with a time window of one year. This time period was based on the fact that all indicators included in the study should be measured at least once a year according to the guidelines.

Ethics statement

The research was conducted under supervision of IQ healthcare, a scientific department focusing on quality improvement in healthcare, and as such forms part of the Radboud University Medical Center in Nijmegen, the Netherlands. LINH is registered with the Dutch Data Protection Authority and data are handled according to the national data protection guidelines. The study was conducted in accordance with the Dutch Law for the Protection of Personal Data (Wet Bescherming Persoonsgegevens)[22] and the Declaration of Helsinki[23]. Based on the Dutch Law for the Protection of Personal Data no medical ethical committee approval was required for this study, therefore the medical ethical committee was not contacted. All data were anonymized before extraction and analysis.

Measures

We included demographic patient characteristics and information on prevalent multi-morbidities in the study. Data were collected on patient age, gender, whether the patient lived in a deprived area and body mass index (BMI). A systematic review showed that out of all conditions that are commonly treated in primary care, CVD, COPD, depression and anxiety most often co-occur with diabetes[13]. Information on multi-morbidity was based on the ICPC codes[24], and was included as a dichotomous variable for each condition. CVD included ischemic heart disease with or without angina, acute myocardial infarction, heart failure, atrial fibrillation, hypertension, transient cerebral ischemia, cerebrovascular disease, stroke

and atherosclerosis. Anxiety and depression included the ICPC codes for feeling depressed, anxiety disorder or anxiety state and depressive disorder. Furthermore, the number of diabetes related consultations per year was included. Whenever the ICPC code for diabetes was recorded in the medical record, the consultation was recorded as diabetes related. Consultations can be provided by the family physician or other staff in the general practice, including a nurse practitioner. In this study, we only included the frequency of the consultations; we did not have information on the caregiver that provided the consultation. We also recorded whether patients received a prescription of lipid lowering medication, anti-diabetic medicine (oral, insulin or both) and/or blood pressure lowering medication.
Outcome measures included 6 dichotomous measures regarding glycosylated hemoglobin (HbA1c), systolic blood pressure (BP) and low density lipoprotein-cholesterol (LDL-C). We included whether the target level (as set in the Dutch guideline[25]) was measured during consultation (yes/no) and, for those cases with a valid measurement, whether it was achieved (yes/no). The most recent valid data for each patient were included in the dataset.

Analysis

Descriptive statistics were used to describe the patient and practice characteristics. A missing data analysis was performed since length and weight (to calculate BMI) were not recorded structurally each year. We performed multiple imputation to complete missing values on BMI data, using 5 imputation sets. A larger number of imputation sets has been recommended. However, since our dataset consisted of a large number of patients, the difficulties of applying a large number of imputation sets (>10) in a reasonable amount of time outweighed the benefits. Main outcomes and predictors were included in the imputation models. We also included practice in the imputation models to account for clustering effects.
To assess which patient factors were associated with clinical processes and outcomes of diabetes care, logistic mixed models were performed on the imputation models in order to obtain pooled results, taking into account the clustering of patients within practices. We ran six separate hierarchical logistic regression models, one for each of the outcome measures (HbA1c, BP and LDL-C measured, HbA1c, BP and LDL-C within target level). Target levels were set at 53 mmol/mol, 140 mm Hg and 2.5 mmol/l respectively. In each model, we included patient age, gender, number of diabetes related consults (above or below median), whether the patient lives in deprived area, BMI and presence of ICPC code related to CVD, COPD and/or depression and anxiety (dichotomous measures). We assessed the discriminative power of the model by calculating the area under the curve in the imputation sets (pooled AUC) and we also calculated pooled fixed effects (OR). In order to assess whether specific factors can be left out from the model without decreasing the effect size, we then reduced each of the 6 models in accordance with the maximum log-likelihood test ($p < 0.10$) and again

calculated pooled fixed effects and pooled AUCs. The multiple imputation and logistic multilevel models were performed in R version 3.03. Descriptive statistics were performed in SPSS version 20.

Table 1. Patient characteristics and overall scores on indicators

Patient characteristics	Study population	Dutch diabetes population*,◊
No. of patients	11718	834100
% of women	49.8%	50.1%
Mean age (SD)	65.8 (13.3)	67.7 (12.6)
Mean Body Mass Index (SD)	29.6 (5.3)	-
% of patients with funds for deprived areas	5.9%	-
Mean no. of consultations per year (SD)	7.6 (6.9)	-
Mean no. of diabetes related consultations per year (SD)	2.8 (3.4)	-
Multi-morbid conditions (% of patients)		
Cardiovascular disease (CVD)	69.8%	-
Chronic obstructive pulmonary disease (COPD)	8.5%	8.4%
Depression and/or anxiety	13.2%	17.6%
Medication prescriptions (% of patients)		
Anti-diabetic medicine		
No medication	31.3%	22%
Oral medication only	52.3%	62%
Oral medication and insulin	10.4%	10%
Insulin only	6.0%	4%
Blood pressure lowering medication	69.3%	-
Lipid lowering medication	64.0%	67.0%
Indicators	**Study population**	**Study multi-disciplinary care**
% of patients with HbA1c value measured	76.7%	91.4%
Mean HbA1c value (SD)	50.8 (11.8)	-
% of patients with HbA1c value below target level (53)	67.3%	68.8%
% of patients with systolic blood pressure measured	79.0%	89.9%
Mean systolic blood pressure value (SD)	140.3 (18.1)	-
% of patients with systolic blood pressure below target level (140)	49.1%	52.3%
% of patients with LDL cholesterol level measured	69.4%	85.3%
Mean LDL cholesterol level (SD)	2.59 (0.93)	-
% of patients with LDL cholesterol level below target level (2.5)	49.6%	53.1%

** Source data: (I) RIVM website[26], (II) publication Zodiac study[27], (III) report on care groups[28], (IV) report on Dutch Accreditation[29]. ◊ Several characteristics of our study population could not be contrasted with the Dutch population or other comparable study populations because no suitable data could be found.*

RESULTS

Study population

11,718 out of 12,498 patients from 59 practices were included in the analyses of process measures. 689 patients were excluded because they were in hospital care instead of primary care, 91 patients were excluded due to age (younger than 18 years). This population is a reasonable representation of all Dutch diabetes patients in primary care, see Table 1.
Subsets of 8,028, 8,936 and 9,257 patients were included in the analysis of outcome measures on LDL cholesterol, HbA1c and systolic blood pressure respectively. The percentage of patients with outcomes below target level was slightly lower than the percentages found in a Dutch multidisciplinary care study; however, in the latter, practices could exclude particular patient groups that are more difficult to treat. Also, the study population scored substantially lower on the registration of the patient outcomes. In Table 2, we report the variation on the process measures across the 59 practices. We found that for each of the process measures, the variation between practices is sufficiently large to show possible effects of patient determinants.

Table 2. Variations in clustered scores (per practice) on process indicators

Indicator outcomes		Study population
HbA1c value measured	Mean	76.0%
	SD (range)	10.2% (50.8% – 96.7%)
Systolic blood pressure measured	Mean	76.9%
	SD (range)	15.0% (7.0% – 96.9%)
LDL cholesterol level measured	Mean	67.0%
	SD (range)	18.3% (2.0% – 93.1%)

In our missing data analysis we found that as expected the main measure with missing data was BMI (3,436 missing, 29.3%). Other variables with missing data included age (1 missing, 0.01%) and whether the patient lives in a deprived area (81 missing, 0.69%). Results from a MCAR test showed that the missing data were not completely at random.

Multilevel logistic regression analysis process measures

Each of the process measure models showed moderate effect sizes, with pooled areas under the curve (AUCs) that varied between 0.66 and 0.76. The closer the AUC value approaches 1, the better the model can distinguish whether processes are performed or not based on the factors included in the model. An AUC of 0.5 would indicate that the model does not perform better than chance. A pooled AUC of 0.76 indicates that for each pair of patients (one with processes performed and one without processes performed), the model could classify these

Table 3. Factors associated with process indicators: measure recorded within study period

Factor*	Complete model		Reduced model		DM consultation only model	
	OR	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
HbA1c						
Age in years	1.03	1.03 – 1.03	1.03	1.02 – 1.03	-	-
Gender male	1.01	0.92 – 1.12	-	-	-	-
Patient in deprived area	1.03	0.78 – 1.36	-	-	-	-
CVD	1.15	1.03 – 1.29	1.17	1.05 – 1.31	-	-
Depression and/or anxiety	1.08	0.93 – 1.25	-	-	-	-
COPD	0.81	0.67 – 0.97	0.81	0.68 – 0.97	-	-
Body mass index in kg/m^2	1.01	1.00 – 1.03	-	-	-	-
DM consultation count >median	14.4	12.3 – 16.7	14.4	12.3 – 16.8	15.4	13.2 – 17.9
AUC	0.76		0.76		0.71	
n	11718		11718		11718	
Systolic blood pressure						
Age in years	1.03	1.03 – 1.04	1.03	1.03 – 1.04	-	-
Gender male	1.04	0.93 – 1.16	-	-	-	-
Patient in deprived area	1.11	0.83 – 1.49	-	-	-	-
CVD	1.75	1.56 – 1.98	1.77	1.57 – 1.99	-	-
Depression and/or anxiety	0.93	0.80 – 1.09	-	-	-	-
COPD	0.72	0.59 – 0.87	0.72	0.59 – 0.87	-	-
Body mass index in kg/m^2	1.01	0.99 – 1.02	-	-	-	-
DM consultation count >median	14.5	12.3 – 17.1	14.5	12.3 – 17.1	15.5	13.1 – 18.2
AUC	0.77		0.77		0.71	
n	11718		11718		11718	
LDL cholesterol						
Age in years	1.02	1.01 – 1.02	1.02	1.01 – 1.02	-	-
Gender male	1.17	1.06 – 1.29	1.16	1.06 – 1.27	-	-
Patient in deprived area	0.87	0.66 – 1.13	-	-	-	-
CVD	1.31	1.18 – 1.46	1.32	1.19 – 1.47	-	-
Depression and/or anxiety	1.03	0.90 – 1.18	-	-	-	-
COPD	0.81	0.69 – 0.95	0.81	0.69 – 0.95	-	-
body mass index in kg/m^2	1.00	0.99 – 1.02	-	-	-	-
DM consultation count >median	5.60	4.99 – 6.28	5.61	5.00 – 6.30	5.92	5.28 – 6.63
AUC	0.70		0.69		0.66	
n	11718		11718		11718	

** Reference categories: female, less than median diabetes related contacts per year, without funds for deprived area, no multi-morbidities.*

correctly in 76% of the cases. The effect sizes could be maintained with a reduced model. In the complete models, HbA1c, systolic blood pressure and LDL cholesterol values were measured more often in older patients, see Table 3. Furthermore, patients with CVD more often had their outcomes measured, with the largest odds for measurement of blood

Table 4. Factors associated with outcome indicators: measure outcome within target level

Factor*	Complete model		Reduced model	
	OR	*95% CI*	*OR*	*95% CI*
HbA1c				
Age in years	1.00	1.00 – 1.01	1.01	1.00 – 1.01
Gender male	0.83	0.75 – 0.91	0.83	0.75 – 0.91
Patient in deprived area	0.84	0.65 – 1.07	-	-
CVD	1.10	0.99 – 1.23	-	-
Depression and/or anxiety	1.19	1.03 – 1.36	1.18	1.03 – 1.36
COPD	0.95	0.81 – 1.12	-	-
Body mass index in kg/m^2	0.97	0.96 – 0.98	0.97	0.96 – 0.98
DM consultation count >median	0.54	0.48 – 0.60	0.54	0.48 – 0.60
AUC	0.59		0.58	
n	8936		8936	
Systolic blood pressure				
Age in years	0.97	0.97 – 0.98	0.97	0.97 – 0.98
Gender male	0.92	0.84 – 1.00	0.92	0.84 – 1.00
Patient in deprived area	0.92	0.71 – 1.18	-	-
CVD	0.54	0.49 – 0.60	0.54	0.49 – 0.60
Depression and/or anxiety	1.07	0.94 – 1.21	-	-
COPD	1.13	0.97 – 1.31	-	-
Body mass index in kg/m^2	0.97	0.96 – 0.98	0.97	0.96 – 0.98
DM consultation count >median	1.04	0.94 – 1.16	-	-
AUC	0.61		0.61	
n	9257		9257	
LDL cholesterol				
Age in years	1.01	1.01 – 1.02	1.01	1.01 – 1.01
Gender male	1.47	1.34 – 1.62	1.45	1.33 – 1.59
Patient in deprived area	1.44	1.09 – 1.89	1.44	1.10 – 1.90
CVD	1.19	1.07 – 1.32	1.20	1.08 – 1.34
Depression and/or anxiety	0.84	0.74 – 0.96	0.84	0.74 – 0.96
COPD	1.03	0.87 – 1.21	-	-
Body mass index in kg/m^2	1.01	1.00 – 1.02	-	-
DM consultation count >median	1.33	1.19 – 1.48	1.33	1.19 – 1.48
AUC	0.57		0.57	
n	8028		8028	

** Reference categories: female, less than median diabetes related contacts per year, without funds for deprived area, no multi-morbidities.*

pressure (odds ratio (OR) 1.75, 95% confidence interval (CI) 1.56 – 1.98). In other words, as a diabetes patient with CVD as multi-morbidity, the odds of having your blood pressure measured were 1.75 times higher than for diabetes patients without CVD. On the other hand, patients with COPD had a lower probability to have their blood pressure measured (OR 0.72,

95% CI 0.59 – 0.87). COPD patients also had a smaller probability of HbA1c and LDL cholesterol measurements compared to patients without COPD. For each of the three process measures there was an association of the frequency of diabetes related consultations and the probability of the value being measured (ORs vary between 5.6 and 14.5). In fact, this factor on its own explained a large part of the effects found in the models (pooled AUC between 0.66 and 0.71) compared to models that included all factors (pooled AUC between 0.70 and 0.76).

Multilevel logistic regression analysis outcome measures

The effect sizes of the outcome measure models were considerably smaller than the process measure models, with pooled AUCs varying from 0.57 to 0.61. This indicates that the model did not perform much better than chance. The probability of achieving the systolic blood pressure target level was lower in older patients, while the probability of an LDL cholesterol level within target was higher in older patients (Table 4). Gender was also associated with outcomes, with lower probabilities of good HbA1c and blood pressure control and a higher probability of good LDL cholesterol control in male patients compared to female patients. Patients with CVD had a higher probability that the HbA1c level was within target level and a lower probability that the blood pressure was within target level. Patients with depression and anxiety also had a higher probability of achieving an HbA1c level within target level. Furthermore, these patients had a lower probability of achieving good LDL cholesterol control. Blood pressure and HbA1c levels were less often below target level in patients with a higher BMI. A relatively high number of diabetes consultations was associated with poorer HbA1c control and better LDL cholesterol control.

DISCUSSION

In this study the associations between several non-modifiable patient characteristics and prevalent multi-morbidities with processes and outcomes of diabetes care in Dutch general practices were explored. We found that the models on processes of care showed moderate effects, suggesting that patient factors had some impact on diabetes care indicators. On the other hand, effects on outcome measures were small. Measurement of the number of diabetes consultations, as a measure of patient health service utilization, was the main contributor to the effects found.

Explanation of findings

Factors incorporated in our study only had small to moderate pooled effects on processes and outcomes of care. This is in line with a study by Marceau et al.[30], who reported that from the 21.1% of variance in diabetes management that could be explained by patient, provider

and organizational factors, only 2.2% could be explained by patient factors. The main contributor was the diabetes consultation frequency. Since in general all patients receive automated invitations for their 3-monthly and yearly consultations, this measure reflects patients' attitudes and compliance. The effects found are in line with other literature that shows that poor adherence to treatment plans (such as failure to come to an appointment) can have negative effects on processes and outcomes of care[31].

Since effect sizes found are moderate to small, it may be debated whether to take these case-mix factors into account in the calculation of scores on population level, either through correction or stratification. The small changes in scores may not weigh up against the troubles of data collection and increased complexity of score interpretation. Furthermore, Nicholl[32] suggests that correction can actually increase the bias in measurements, especially when interactions between factors are not taken into account properly. However, it is possible that other patient factors than the factors in this study are relevant for diabetes care processes and outcomes, or that the examined factors are relevant for health outcomes that were not examined in this study. Also, for the evaluation of individual patients it is important to be aware of the possible risk factors, since they may have large influence in a single patient. Furthermore, when the total number of diabetes patients in a practice is relatively small, one should account for a larger variation in the score[33]. In such case, we suggest that accounting for a larger error margin is better than correcting for multiple factors, since this will only increase the bias due to the lower precision of the data.

Results on measurement of outcomes are in line with previous literature that shows that the influence of a multi-morbidity depends on whether the conditions are strongly related to each other[20]. We found that in patients with COPD, a condition that is less strongly related to diabetes, processes of care were less often according to the guidelines. It is possible that during consultations with COPD patients, priority is given to address those issues that cause the highest burden on the patient quality of life the most. These issues are more likely to be COPD related rather than diabetes related. On the other hand, patients with a strongly related condition, CVD, were more likely to have their blood pressure measured. Results are similar to a Dutch study by Nouwens et al., in which better cardiovascular risk treatment in diabetes patients but not in COPD patients was reported[16]. Our findings are also in line with the study by Voorham et al.[21]. They looked at the association between multi-morbidity and treatment intensification. In their study, new occurrences of diabetes-related multi-morbidities such as cardiovascular disease were related with better risk factor treatment, whereas no effects were found for non-diabetes related multi-morbidities.

Our results show that patients with depression and anxiety have a higher probability to have good HbA1c control than others, whereas the probability is lowered that they have good cholesterol control. Although literature shows that depression is associated with poorer

medication adherence[34], evidence regarding the effects on HbA1c, blood pressure and cholesterol control is inconclusive[35,36]. Possibly patients with anxiety are more sensitive for side-effects of lipid lowering medication, resulting in poor cholesterol control.

Strengths and limitations

In this observational study, there were several factors associated with diabetes care that we could not take into account. These include hereditary factors and lifestyle factors, which have known associations with development of diabetes[37]. We also did not include duration and severity of the illness in our study, which is an important limitation. Furthermore, medication compliance may have varied between patients, which may have also influenced the scores. Other studies show that longer diabetes duration is associated with more diabetes complications such as retinopathy[38] and coronary heart disease[39]. These factors may explain part of the variation that we could not account for with our model. Poor outcomes of care may be accounted for by factors that we did not include in our dataset, including presence of pancreatic illnesses or issues with medication tolerance. However, these factors only regard small parts of the population of diabetic patients seen by the family physician. Therefore, we expect that this has not had a large influence on our study.

Our study was performed on a large dataset that formed an adequate reflection of the diabetes patients treated in primary care. There were several limitations to the use of this dataset. We could only analyze the outcomes of care for those patients that had a measurement of the outcome, i.e. scored positively on the process measure. This can cause a bias in our study. It may be the case that physicians are more likely to perform measurements when they expect that patients need treatment; leading to a slightly lower percentage of patients with a measurement within target level compared to those without a measurement. On the other hand, there may have been an underrepresentation of patients that show higher noncompliance to treatment. It is to be expected that outcomes are not within target level for this group of patients, which may cause a bias in the opposite direction. However, we argue that this bias also occurs in routine care. Since our aim was to develop a model of determinants that approaches everyday care, the large study population that we used was valid for its purpose.

Since data were collected through an automated extraction of medical records, it is likely that there is an effect of poor registration. This is reflected by the large difference in measurement of outcomes in our study population compared to the report on multidisciplinary care. In the latter study, the practices collected data from their registrations themselves, which meant that they could retrieve misplaced information more easily. The registration bias in our study may have led to a bias in our study population. Although poor registration may have led to

lower scores on processes of care, in general our population was reasonably well controlled. This may have decreased the likelihood of finding associations in this study.

Conclusions

Several non-modifiable patient factors could be associated with processes and outcomes of diabetes care. The main contributor to improved processes of care was the frequency of diabetes consultations. However, especially regarding the outcomes of care, associations were small. Therefore, our results do not support the need for case-mix correction or stratification suggested in previous literature. Accounting for a certain margin of error around the scores without further correction may be more beneficial for users in practice.

REFERENCES

1. Renders CM, Valk GD, Griffin SJ, Wagner EH, van Eijk JT, et al. Interventions to improve the management of diabetes in primary care, outpatient, and community settings: a systematic review. Diabetes Care 2001; 24(10):1821-33.
2. Beaton SJ, Nag SS, Gunter MJ, Gleeson JM, Sajjan SS, et al. Adequacy of glycemic, lipid, and blood pressure management for patients with diabetes in a managed care setting. Diabetes Care 2004;27(3):694-8.
3. Skyler JS, Bergenstal R, Bonow RO, Buse J, Deedwania P, et al. Intensive glycemic control and the prevention of cardiovascular events: implications of the ACCORD, ADVANCE, and VA Diabetes Trials: a position statement of the American Diabetes Association and a Scientific Statement of the American College of Cardiology Foundation and the American Heart Association. J Am Coll Cardiol 2009; 53(3):298-304.
4. Malik S, Lopez V, Chen R, Wu W, Wong ND. Undertreatment of cardiovascular risk factors among persons with diabetes in the United States. Diabetes Res Clin Pract 2007;77(1):126-33.
5. Schaars CF, Denig P, Kasje WN, Stewart RE, Wolffenbuttel BH, et al. Physician, organizational, and patient factors associated with suboptimal blood pressure management in type 2 diabetic patients in primary care. Diabetes care 2004;27(1):123-8.
6. Campbell JA, Walker RJ, Smalls BL, Egede LE. Glucose control in diabetes: the impact of racial differences on monitoring and outcomes. Endocrine 2012;42(3):471-82.
7. Strom Williams JL, Lynch CP, Winchester R, Thomas L, Keith B, et al. Gender differences in composite control of cardiovascular risk factors among patients with type 2 diabetes. Diabetes Technol Ther 2014;16(7):421-7.
8. Chin MH. Risk-adjusted quality of care rating for diabetes: ready for prime time? Diabetes care 2000;23(7):884-6.
9. Zaslavsky AM, Hochheimer JN, Schneider EC, Cleary PD, Seidman JJ, et al. Impact of sociodemographic case mix on the HEDIS measures of health plan quality. Med Care 2000;38(10):981-92.
10. Paddison C, Elliott M, Parker R, Staetsky L, Lyratzopoulos G, et al. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. BMJ Qual Saf 2012;21(8):634-40.
11. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: what's in a name? A review of literature. Eur J Gen Pract 1996;2(2):65-70.
12. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. Eur J Gen Pract 2008;14 Suppl 1:28-32.
13. Sinnige J, Braspenning J, Schellevis F, Stirbu-Wagner I, Westert G, et al. The prevalence of disease clusters in older adults with multiple chronic diseases – a systematic literature review. Plos One 2013;8(11):e79641.
14. Struijs JN, Baan CA, Schellevis FG, Westert GP, van den Bos GA. Comorbidity in patients with diabetes mellitus: impact on medical health care utilization. BMC Health Serv Res 2006; 6:84.
15. Woodard LD, Urech T, Landrum CR, Wang D, Petersen LA. Impact of comorbidity type on measures of quality for diabetes care. Med Care 2011;49(6):605-10.
16. Nouwens E, van Lieshout J, Wensing M. Comorbidity complicates cardiovascular treatment: is diabetes the exception? Neth J Med 2012;70(7):298-305.
17. Bae S, Rosenthal MB. Patients with multiple chronic conditions do not receive lower quality of preventive care. J Gen Intern Med 2008;23(12):1933-9.
18. Hughes LD, McMurdo ME, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age Ageing 2013;42(1):62-9.
19. Kerr EA, Heisler M, Krein SL, Kabeto M, Langa KM, et al. Beyond comorbidity counts: how do comorbidity type and severity influence diabetes patients' treatment priorities and self-management? J Gen Intern Med 2007; 22(12):1635-40.
20. Turner BJ, Hollenbeak CS, Weiner M, Ten Have T, Tang SS. Effect of unrelated comorbid

conditions on hypertension management. Ann Intern Med 2008;148(8):578-86.
21. Voorham J, Haaijer-Ruskamp FM, Wolffenbuttel BH, de Zeeuw D, Stolk RP, et al. Differential effects of comorbidity on antihypertensive and glucose-regulating treatment in diabetes mellitus: a cohort study. Plos One 2012;7(6):e38707.
22. website wet bescherming persoonsgegevens. http://wetten.overheid.nl/BWBR0011468.
23. website WMA Declaration of Helsinki – Ethical principles for medical research involving human subjects. http://www.wma.net/en/30publications/10policies/b3/.
24. Bentsen BG. International classification of primary care. Scand J Prim Health Care 1986; 4(1):43-50.
25. NHG-standaard diabetes mellitus type 2 (derde herziening). Huisarts Wet 2013;56(10): 512-25.
26. website RIVM, nationaal kompas volksgezondheid. http://www.nationaalkompas.nl/ gezondheid-en-ziekte/ziekten-en-aandoeningen.
27. Ubink-Veltmaat LJ, Bilo HJ, Groenier KH, Houweling ST, Rischen RO, et al. Prevalence, incidence and mortality of type 2 diabetes mellitus revisited: a prospective population-based study in The Netherlands (ZODIAC-1). Eur J Epidemiol 2003;18(8):793-800.
28. Landelijke Organisatie voor Ketenzorg. Transparante ketenzorg Diabetes mellitus 2010. Utrecht: LHV/VHN/LVG; 2012.
29. van Doorn-Klomberg AL, Braspenning JC, Bouma M. Rapportage NHG-Praktijkaccreditering: gegevens over het medisch handelen in 2012. Nijmegen: IQ healthcare, Radboud umc; 2013.
30. Marceau L, McKinlay J, Shackelton R, Link C. The relative contribution of patient, provider and organizational influences to the appropriate diagnosis and management of diabetes mellitus. J Eval Clin Pract 2011;17(6):1122-8.
31. Nam S, Chesla C, Stotts NA, Kroon L, Janson SL. Barriers to diabetes management: patient and provider factors. Diabetes Res Clin Pract 2011;93(1):1-9.
32. Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. J Epidemiol Community Health 2007;61(11):1010-3.
33. van Doorn-Klomberg AL, Braspenning JC, Feskens RC, Bouma M, Campbell SM, et al. Precision of individual and composite performance scores: the ideal number of indicators in an indicator set. Med Care 2013; 51(1):115-21.
34. Gonzalez JS, Peyrot M, McCarl LA, Collins EM, Serpa L, et al. Depression and diabetes treatment nonadherence: a meta-analysis. Diabetes Care 2008;31(12):2398-403.
35. Heckbert SR, Rutter CM, Oliver M, Williams LH, Ciechanowski P, et al. Depression in relation to long-term control of glycemia, blood pressure, and lipids in patients with diabetes. J Gen Intern Med 2010;25(6):524-9.
36. Lustman PJ, Clouse RE. Depression in diabetic patients: the relationship between mood and glycemic control. J Diabetes Complications 2005;19(2):113-22.
37. Qi L, Hu FB, Hu G. Genes, environment, and interactions in prevention of type 2 diabetes: a focus on physical activity and lifestyle changes. Curr Mol Med 2008;8(6):519-32.
38. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 2012;35(3):556-64.
39. Fox CS, Sullivan L, D'Agostino RB Sr, Wilson PW. The significant effect of diabetes duration on coronary heart disease mortality: the Framingham Heart Study. Diabetes Care 2004; 27(3):704-8.

7

Organizational determinants of high-quality routine diabetes care

Arna van Doorn - Klomberg
Jozé Braspenning
René Wolters
Margriet Bouma
Wim de Grauw
Michel Wensing

Scandinavian Journal of Primary Health Care 2014; 32(3): 124-31.

ABSTRACT

Background Randomized trials showed that changes in healthcare organization improved diabetes care. This study aimed to identify which organizational determinants were associated with patient outcomes in routine diabetes care.

Methods We performed an observational study, in which multilevel regression analyses were applied to examine the impact of 12 organizational determinants on diabetes care as separate measures and as a composite score. We included 11,751 patients with diabetes in 354 general practices in the Netherlands. Main outcome measures included patients' recorded glycated hemoglobin (HbA1c), systolic blood pressure, and serum cholesterol levels.

Results A higher score on the composite measure of organizational determinants was associated with better control of systolic blood pressure (p = 0.017). No effects on HbA1c or cholesterol levels were found. Exploration of specific organizational factors found significant impact of use of electronic medical records on HbA1c (OR = 1.80, 95% CI 1.12 – 2.88), availability of patient leaflets on systolic blood pressure control (OR = 2.59, 95% CI 1.06 – 6.35), and number of hours nurse education on cholesterol control (OR = 2.51, 95% CI 1.02 – 6.15).

Conclusions In routine primary care, it was found that favorable healthcare organization was associated with a number of intermediate outcomes in diabetes care. This finding lends support to the findings of trials on organizational changes in diabetes care. Notably, the composite measure of organizational determinants had most impact.

BACKGROUND

Improving diabetes care has been on the agenda for several decades in many countries. In their systematic review of 142 trials, Tricco et al. examined the effects of various organizational changes on glycated hemoglobin (HbA1c), blood pressure, and serum cholesterol levels[1]. They identified 12 target areas (Table 1) in three domains: the local healthcare system, the healthcare providers, and the patients.

The results suggested that in particular the combination of intervention components targeting the health system and patient-mediated interventions contributed to better outcomes. Since the amount of different intervention components varied among the studies and the relationship between components was often not investigated, considerable uncertainty remained associated with these findings. Similar results were found in a recent review by Stellefson et al.[2]. This study showed that interventions related to each of the six Chronic Care Model (CCM) components[3,4], 'community resources and policies', 'healthcare organization', 'self-management support', 'delivery system design', 'decision support', and 'clinical information systems', contributed to quality improvement. However, there was not one particular component that proved to be a key component to achieve improvements. They also suggest that a multifaceted program could facilitate better implementation. Most evaluative studies are randomized controlled trials, which involve optimized support for achieving change. Although this offers obvious advantages regarding risk of bias, there is inconclusive evidence on whether effects can be replicated in daily practice[5,6].

In this contribution, we focus on the impact of organizational determinants on (intermediate) patient outcomes in routine diabetes care for type 2 diabetic patients. In the Netherlands, this is largely provided in primary care. We aimed to identify organizational determinants of the following outcomes: HbA1c, blood pressure, and cholesterol levels. As the quality of diabetes care generally has improved over the last decades[7-11], we wondered whether organizational determinants identified in trials had been implemented and whether they contributed to increased control in routine diabetes care in the Netherlands.

Since several studies suggested that a combination of determinants rather than individual determinants contribute most to quality improvement[1,2], we also examined the impact of a composite measure of organizational factors.

Table 1. Measures included in the study related to the target areas for improvement strategies

Tricco taxonomy		Included measure
Health systems		
Team changes	1	Nurse practitioner volume: the amount of FTE (full-time equivalent) NP (nurse practitioner) per 1,000 patients
Continuous QI		*Continuous quality improvement in the form of written plans with plan-do-study-act cycles was present in all practices since this was required in the Dutch Accreditation program*
	2	Annual report: dichotomous measure of whether the practice evaluates quality in a written report (for internal use) at least once a year
	3	Complaints procedure: percentage of patients aware of the complaints procedure in the practice
Electronic medical records	4	Use of Electronic Medical Records (EMR), sum score of 7 dichotomous items: a) Family physicians (FPs) always use the EMR to create prescriptions b) Incoming lab results are processed automatically c) Hospital referrals are completely created in the EMR d) Referrals to other disciplines (e.g. physiotherapy) are completely created in the EMR e) Application forms for diagnostic procedures are generated in the EMR f) Contra-indications and intolerances are systematically recorded in the EMR g) FPs have the support of an electronic referral system during visiting hours
Case management	5	Diabetic clinic available in the practice (dichotomous)
Facilitated relay of information to clinicians	6	Consultation with partners, sum score of five dichotomous items: a) Practice has regular consultations with local district nurses b) Practice has regular consultations with local physiotherapists c) Practice has regular consultations with local dieticians d) Practice has regular consultations with local pharmacists e) Practice has regular consultations with local social workers
	7	Collaboration with partners, sum score of 2 dichotomous items: a) Practice collaborates with local physiotherapists b) Practice collaborates with local social workers
Healthcare providers		
Financial incentives		-
Clinician education	8	FP education: dichotomous measure, amount of accredited education less than 50 hours per year or exactly/more than 50 hours. The cutoff point was based on the approximate median score in our dataset.
	9	Nurse education: dichotomous measure, amount of education less than 15 hours per year or exactly/more than 15 hours. The cutoff point was based on the approximate median score in our dataset.
	10	Electronic guidelines are available in every treatment room (dichotomous)
Clinician reminders		-
Audit and feedback		*The Dutch Accreditation program was an audit and feedback system in itself, in which all practices participated*

Table 1. (continued)

Tricco taxonomy		Included measure
Patients		
Education of patients	11	Patient leaflets, composite score of two dichotomous items: a) Patient leaflets regarding cardiovascular diseases are available in the practice b) All patient leaflets are kept in an area that is clearly visible and accessible for patients
Promotion of self-management		-
Reminder systems	12	Patient reminder system available in the practice (dichotomous)

METHODS

Study design and study population

We analyzed electronic patient record data from 362 general practices that participated for the first time in the Dutch Accreditation program between 2006 and 2009 (Box 1). These practices manually extracted information from a random sample of 40 medical records, using a structured protocol and data-extraction tool to ensure the selected sample was unbiased. Data were collected retrospectively during the preparatory phase of the accreditation process. The practice was instructed to select records on only those patients that were diagnosed with diabetes and had a family physician (FP) as main diabetes care provider. In effect, practices mainly included diabetes type 2 patients.

Box 1. Dutch Accreditation program for general practices

Since 2005, general practices can voluntarily take part in the Dutch Accreditation program. The preparatory phase of this program consists of the collection of data on practice management and patient care. The measurement instruments used are previously validated questionnaires such as the 'VIP', a visitation instrument for practice organization and the 'Europep' that measures patient experiences[12,13]. The questionnaires are filled in by family physicians (FPs), nurses and patients. There are also questionnaires for a trained consultant that pre-audits the practice. Clinical performance is measured with the use of patient information that is extracted from electronic medical records; the FP or nurse extracts the information either automatically or manually with an extraction form.

When all data are collected and submitted through an online questionnaire system, the practice receives a report that includes information on its own performance and the performance of other practices as benchmarks. This information helps to identify which areas could be improved upon. The FPs then write improvement plans with a plan-do-study-act cycle. The first audit is carried out after the approval of these plans to confirm adequate participation and to grant accreditation. After this audit a three-year accreditation cycle starts. At the end of each year the practice staff evaluates whether the objectives of improvement plans are met and writes new improvement plans for the following year. The prolongation of the accreditation depends on this process. Accreditation is not based on the actual quality of care itself but rather on the quality of the improvement initiatives according to a structured program. After three years, a new cycle starts with the data collection phase. In our current study, we have excluded these repeated measurements; we have only included data from the first cycle.

Measures

Our outcome measures were recorded HbA1c, systolic blood pressure, and total cholesterol values. Organizational determinants related to high-quality diabetes care, derived from a review of trials[1], were linked to specific items from validated questionnaires[12]. FPs, other practice staff, patients, and trained observers filled in these questionnaires as part of the Dutch Accreditation program. We composed 12 measures, some with multiple items (see Table 1), and their composite score (see Appendix 1). Furthermore, we gathered data from these questionnaires on practice type, size and urbanization, the number of diabetes patients per 1,000 patients, and the volume of FP per 1,000 patients (expressed as a full-time equivalent, FTE). Patients' age and gender were extracted from the patients' medical record, as well as the year of data collection (2006/2007/2008/2009). Patients' age was considered as a proxy for comorbidity.

Analysis

We used means and percentages as appropriate to summarize patient and practice characteristics and establish whether determinants were implemented in our study population. Correlations of all measures (Pearson's correlation coefficient or Spearman's correlation coefficient as appropriate) were calculated to check for multicollinearity; the cutoff point for exclusion was set at $r > 0.15$. We performed separate multilevel regression analyses on each of the three outcome measures: HbA1c, total cholesterol, and systolic blood pressure in SPSS (version 20). We examined the effect of the composite score of the 12 determinants, while controlling for patient characteristics (age and gender) and practice characteristics (practice type, year of data collection). Other practice characteristics were not included in the model because of correlations above $r = 0.15$ with practice type. In order to explore which underlying determinants of the composite score were related to outcomes, we repeated the multilevel regression analyses on each of the 12 determinants in Table 1 separately. As effects may be smaller in patient populations with higher baseline control[1], we also performed logistic regression analyses while comparing the extremes, that is the practices that performed in the highest and lowest quartiles.

RESULTS

Study population

Patients who were mainly under specialist care during the study period were excluded. Data on 11,751 diabetes patients from 354 practices remained (mean value of 33.2 patients per practice, minimally 10 patients per practice). The study population contained slightly more large practices in urban regions compared with the Dutch population (Table 2). Practices that

offer vocational training were overrepresented, as it was already known that the Accreditation program would become mandatory for them in the ensuing years.

Table 2. Description of participating practices and scores on determinants

Description	Study population (2006 – 2009)	Dutch population (2008)[14-16]
Practice characteristics	*% or value*[†]	*% or value*[†]
Practice type* (% of practices)		
Single handed practice	20.2%	42.3%
Non - single handed, not in health center	57.6%	
Practice with 2 FPs	23.6%	31.5%
Group practice (>2 FPs)	13.8%	26.1%
Other practice type	20.2%	
Practice within primary health center	22.2%	
Practice size: no. of patients	4961	3888
Number of FTE[◊] FP per 1000 patients	0.43	0.43
Practice location[‡]		
Rural	12.6%	12.2%
Partly urban	38.2%	41.1%
Urban	49.2%	46.7%
At least one of the FPs in the practice provides vocational training	48%	30%
Total % of FPs that provide vocational training	31%	19%
No. of diabetic patients per 1000 patients	43	43
Year of participation		
2005/2006	22.9%	
2007	42.2%	
2008/2009	34.8%	
Scores on determinants	*% or mean (SD)*	
Volume of NP (FTE per 1000 patients)[^]	0.14 (0.06)	
Availability of annual report	48%	
% of patients familiar with the complaints procedure	51.1 (12.5)	
Sum score EMR[#] (score between 0 and 7)	5.38 (1.03)	
Availability of diabetic clinic	88%	
Sum score consultation partners (score between 0 and 5)	3.77 (1.18)	
Sum score collaboration partners (score between 0 and 2)	0.46 (0.69)	
FP education (hours per year)	51 (12.9)	
Nurse education (hours per year)	17 (17.2)	
Availability of guidelines (score between 0 and 1)	0.95 (0.17)	
Sum score patient leaflets (score between 0 and 2)	1.68 (0.56)	
Availability of patient reminder system	78%	

Table 2. (continued)

Description	Study population (2006 – 2009)	GIANTT 2007[17]	
Patient characteristics	*% or mean (SD)*	*% or mean (SD)*	
Patient age	65.9 (12.1)	66.6 (12.3)	
Patient gender, male	51%	48%	
HbA1c value *% / mmol/mol*	6.8 (1.0) / 51 (10.9)	6.9 (1.0) / 52 (10.9)	
Total cholesterol value *mmol/l*	4.7 (1.1)	4.4 (1.1)	
Systolic blood pressure value *mm Hg*	140 (18)	142 (20)	
HbA1c within target value (7.0)	67.4%	61%	
Total cholesterol within target value (5.0)	63.2%		
Systolic blood pressure within target value (150)	70.9%		
Within 140	46.9%	44%	
Within 160	84.0%	87%	
Aggregated average and percentile scores on practice level	*25th percentile*	*Average*	*75th percentile*
HbA1c value *% / mmol/mol*	7.0 / 53	6.8 / 51	6.5 / 48
Total cholesterol value *mmol/l*	4.9	4.7	4.5
Systolic blood pressure value *mm Hg*	144	140	136

** The figures based on the Dutch population distinguish between solo, duo and group practices. In our study there were also the options 'part of health center', which can be either solo, duo or group, and 'other practice type', which consists mostly of duo or group practices within a cooperation construction other than a health center. ◊ FTE = Full-time equivalent. ‡ A region was defined as urban when the number of addresses per km² exceeded 1,500; partly urban regions had 500 - 1.500 addresses per km²; rural regions had less than 500 addresses per km². ^ NP = Nurse practitioner. # EMR = Electronic medical records. † Because the percentages were rounded off, they do not always add up to exactly 100%.*

Measures

Table 2 gives the mean scores and standard deviations (SD) or percentages (in case of dichotomous measures) on the measures included in the study. The percentage of patients with a value below the target for HbA1c, cholesterol, and blood pressure was relatively high (between 63% and 71%). However, 68.4% of all patients had at least one intermediate outcome that was above the target, i.e. diabetes control was not according to guidelines on all aspects.

Organizational determinants

The included practices had high scores on the organizational determinants, indicating favorable conditions for high-quality diabetes care (see Table 2). Determinants that were implemented on a broad scale included the diabetic clinic, a patient reminder system, and guideline availability. Also, in most practices (89%) a nurse practitioner (NP) was part of the practice team. More possibilities for further implementation were found on collaboration with partners and the familiarity of patients with the complaints procedure.

Table 3. Multilevel models (cluster measure = practice, data on patient level) on continuous outcomes of HbA1c, cholesterol and blood pressure for determinant composite score

Parameter*	HbA1c (%), n = 7281		Total cholesterol (mmol/l), n = 7122		Systolic blood pressure (mm Hg), n = 7320	
	Estimate (95% CI)	*p*	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*
Intercept	6.77 (6.53 – 7.01)	.000	4.54 (4.30 – 4.77)	.000	118.8 (114.9 – 122.7)	.000
Sum score determinants	.008 (-.018 – .035)	.546	.005 (-.020 – .030)	.698	-.50 (-.91 – -.09)	.017
Patient age	-.002 (-.004 – -.000)	.049	-.001 (-.003 – .001)	.419	.34 (.31 – .38)	.000
Patient gender	.028 (-.017 – .074)	.221	-.354 (-.402 – -.306)	.000	-.53 (-1.33 – .28)	.199
Year data collection						
2007	.071 (-.042 – .185)	.217	-.014 (-.122 – .095)	.804	2.39 (.61 – 4.18)	.009
2008	.079 (-.038 – .196)	.187	-.170 (-.282 – -.059)	.003	2.30 (.46 – 4.13)	.014
2009	.149 (-.028 – .325)	.098	-.091 (-.258 – .076)	.284	1.68 (-1.08 – 4.43)	.231
Practice type						
Single handed	-.020 (-.134 – .094)	.729	-.026 (-.134 – .083)	.638	-1.39 (-3.18 – .39)	.125
Health center	.076 (-.042 – .194)	.206	.149 (.037 – .262)	.010	.02 (-1.82 – 1.87)	.981

** Reference categories: patient gender male, data collection in 2006, practice with more than 1 FP that is not part of a health center.*

Impact of composite organizational measure

Table 3 shows that systolic blood pressure levels were lower in practices with a higher determinant sum score (p = 0.017). The additional implementation of one determinant led to a decrease of 0.5 in the systolic blood pressure, with a maximum estimated decrease of six points. No effects were found regarding HbA1c or cholesterol levels.

Impacts of organizational items

Of 36 possibilities (12 items x 3 outcomes) we found three significant effects. HbA1c levels were lower in practices that made more use of their electronic medical records (B = -0.088, p < 0.000). HbA1c levels were higher in practices with a diabetic clinic (B = 0.327, p < 0.000), and practices that wrote an annual report (B = 0.103, p = 0.032). None of the 12 measures was found to have an influence on total cholesterol or systolic blood pressure, see Appendix 2.

Comparison highest and lowest quartile practices

Of the 34 measured effects, three effects were significant. Practices that made more use of their electronic medical records (increase of one on a scale of seven) were more likely to score in the highest quartile (OR = 1.8), see Table 4. If the amount of nurse education was relatively high (above the median), there was a higher likelihood that the practice performed within the

highest quartile on total cholesterol (OR = 2.51). Regarding the average systolic blood pressure, better availability of patient leaflets increased the odds (OR = 2.59) that practices comprised part of the 'best practices'.

Table 4. Logistic regression comparison practices in highest and lowest quartile, corrected for patient age, gender, practice type and year of data collection

Parameter	HbA1c, n = 109		Total cholesterol, n = 120		Systolic blood pressure, n = 122	
	OR (95% CI)	*p*	*OR (95% CI)*	*p*	*OR (95% CI)*	*p*
Volume of NP*	.031 (.00 – 11.94)	.253	6.25 (.00 – 18334.95)	.653	29.62 (.03 – 33804.60)	.345
Annual report	.40 (.14 – 1.13)	.083	.68 (.25 – 1.83)	.442	.66 (.22 – 1.94)	.447
Complaints policy	1.02 (.98 – 1.07)	.294	.99 (.95 – 1.03)	.481	.98 (.94 – 1.03)	.403
Use of EMR◊	1.80 (1.12 – 2.88)	.014	.97 (.64 – 1.48)	.892	1.65 (.95 – 2.85)	.074
Diabetic clinic	‡	‡	.55 (.08 – 4.01)	.554	.37 (.06 – 2.07)	.255
Consultation	1.40 (.91 – 2.16)	.126	1.00 (.70 – 1.43)	.992	1.20 (.81 – 1.78)	.359
Collaboration	1.19 (.57 – 2.51)	.643	1.27 (.62 – 2.63)	.512	1.40 (.68 – 2.88)	.363
FP education	.75 (.31 – 1.82)	.519	.80 (.32 – 1.99)	.636	1.62 (.65 – 4.05)	.301
Nurse education	1.13 (.45 – 2.82)	.800	2.51 (1.02 – 6.15)	.045	1.15 (.43 – 3.11)	.778
Guidelines	‡	‡	1.01 (.06 – 17.40)	.997	.34 (.01 – 14.76)	.571
Patient leaflets	.40 (.16 – 1.01)	.051	1.11 (.41 – 3.04)	.833	2.59 (1.06 – 6.35)	.037
Reminder system	1.03 (.30 – 3.57)	.958	.66 (.23 – 1.89)	.443	1.39 (.44 – 4.41)	.579

** NP = Nurse practitioner. ◊ EMR = electronic medical records. ‡ The determinants diabetic clinic and guidelines could not be included in the analysis regarding HbA1c because these variables were a constant in one of the quartile groups.*

DISCUSSION

In this study, targeted organizational determinants were broadly implemented in routine primary care. A higher score on the composite measure of healthcare organization was associated with better systolic blood pressure control. Exploration of specific organizational factors identified only three significant effects in 36 combinations. Comparison of highest and lowest quartiles yielded similar results (three significant effects out of 34). These results from daily practice suggest that implementation of a combination of organizational determinants (rather than individual organizational items) is crucial for high-quality diabetes care[1,2,18].

Strengths and limitations

The current study monitored ongoing care in an average Dutch practice, as opposed to a controlled trial setting where active changes are made to practice management. Most organizational determinants were broadly implemented in routine care, reflecting a

longstanding process of improving diabetes care in the Dutch primary care setting. Perhaps the voluntariness of participation in this study contributed to this fact.
Outcomes were similar in the Dutch GIANTT study[17], in which all diabetic patients from general practices were included. This suggests that the selected patient samples formed an adequate reflection of the practice population. The relatively well-controlled population and the small amount of variation within the organizational measures decreased the potential impact of organizational items[19]. Although other individual studies have used similar items to operationalize constructs[20], there were some constructs that were based solely on one dichotomous item, which could have affected the representation of the model.

Explanation of findings

Our observational study showed that, similar to trials, a multifaceted approach to improve the quality of care involving a combination of organizational interventions can be expected to achieve larger effects than single interventions[1,2]. Ose et al. found that a practice management program had a positive effect on quality of life outcomes[21]. Coleman et al. evaluated the effects of implementation of the Chronic Care Model (CCM) and concluded that an integrated approach was positively associated with improvements in organization and outcomes of care[22]. Compared with other studies, effect sizes in our study were moderate, as was their clinical relevance. Nevertheless, the differences between practices with highest and lowest quartile organizational performance were substantial. Further research is required to unravel whether multiple favorable organizational determinants imply additive or multiplicative effects. On average, the outcome measures in our study met the standard levels as described in the diabetes guidelines, but the composite measure for organizational determinants still showed a significant effect regarding blood pressure. This is probably due to the fact that blood pressure offered the largest room for improvement[17,23,24].
Regarding the individual organizational determinants we found few and small effects. There was a positive effect of the use of the Electronic Medical Records (EMR) on diabetes care, which is consistent with the research evidence from the trials[25]. The positive relation between the availability of patient leaflets and blood pressure has been suggested in other studies as well[26]. Less clear was the unexpected negative association of the availability of a diabetic clinic and an annual report with HbA1c. However, in the Netherlands both features are related to large practices, which in turn have a negative association with quality of care[27,28]. The data supported this argument, as practice size was somewhat larger for practices with an annual report (5,279 versus 4,526 patients) as well as for practices with a diabetes clinic (5,086 versus 3,910 patients).
In our study we used the target values as described in the diabetes guidelines at the time of measurement. However, in daily practice, these are influenced by patient values and

preferences and should not be handled strictly. It is likely that FPs might have deviated from the guidelines when they thought strict treatment was not preferred. For instance, in our dataset the percentage of people with a blood pressure above target was higher for the elderly (70 and above) than the younger population (39% versus 30%).

Conclusions

In line with previous research a combination of determinants of practice organization was more strongly related to meeting the targets on diabetes management than a single determinant[1,2,22]. On average the targets for the management of diabetes care as described in the diabetes guidelines were met, but improvement on intermediate diabetes outcomes could still be reached by introducing more structured practice organization.

REFERENCES

1. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet 2012;379(9833):2252-61.
2. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. Prev Chronic Dis 2013;10:E26.
3. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-44.
4. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002;288(14):1775-9.
5. Cardona-Morrell M, Rychetnik L, Morrell SL, Espinel PT, Bauman A. Reduction of diabetes risk in routine clinical practice: are physical activity and nutrition interventions feasible and are the outcomes from reference trials replicable? A systematic review and meta-analysis. BMC Public Health 2010;10:653.
6. Seenan P, Long M, Langhorne P. Stroke units in their natural habitat: systematic review of observational studies. Stroke 2007;38(6):1886-92.
7. Campbell S, Steiner A, Robison J, Webb D, Raven A, et al. Is the quality of care in general medical practice improving? Results of a longitudinal observational study. Br J Gen Pract 2003;53(489):298-304.
8. Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. BMJ Qual Saf 2013;22(1): 53-64.
9. McWilliams JM, Meara E, Zaslavsky AM, Ayanian JZ. Differences in control of cardiovascular disease and diabetes by race, ethnicity, and education: U.S. trends from 1999 to 2006 and effects of medicare coverage. Ann Intern Med 2009;150(8):505-15.
10. Mundet X, Cano F, Mata-Cases M, Roura P, Franch J, et al. Trends in chronic complications of type 2 diabetic patients from Spanish primary health care centres (GEDAPS study): ten year-implementation of St. Vincent recommendations. Prim Care Diabetes 2012; 6(1):11-8.
11. Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Trends in the quality of care and racial disparities in Medicare managed care. N Engl J Med 2005;353(7):692-700.
12. van den Hombergh P, Grol R, van den Hoogen HJ, van den Bosch WJ. Assessment of management in general practice: validation of a practice visit method. Br J Gen Pract 1998; 48(436):1743-50.
13. Grol R, Wensing M, Mainz J, Jung HP, Ferreira P, et al. Patients in Europe evaluate general practice care: an international comparison. Br J Gen Pract 2000;50(460):882-7.
14. website statline database, CBS Nederland, 2008. http://www.cbs.nl/en-GB/menu/home/default.htm.
15. Hingstman L, Kenens RJ. Cijfers uit de registratie van huisartsen – peiling 2008. Utrecht: NIVEL; 2009.
16. SBOH: data extractie database. Utrecht: SBOH; 2013.
17. Voorham J, Haaijer-Ruskamp FM, van der Meer K, de Zeeuw D, Wolffenbuttel BH, et al. Kwaliteit van de behandeling van type 2-diabetes: resultaten van het GIANTT-project 2004-2007. Ned Tijdschr Geneeskd 2010; 154:A775.
18. Ludt S, Campbell SM, Petek D, Rochon J, Szecsenyi J, et al. Which practice characteristics are associated with the quality of cardiovascular disease prevention in European primary care? Implement Sci 2013;8:27.
19. Fung V, Schmittdiel JA, Fireman B, Meer A, Thomas S, et al. Meaningful variation in performance: a systematic literature review. Med Care 2010;48(2):140-8.
20. Shojania KG, Ranji SR, McDonald KM, Grimshaw JM, Sundaram V, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. JAMA 2006;296(4):427-40.
21. Ose D, Wensing M, Szecsenyi J, Joos S, Hermann K, et al. Impact of primary care-based disease management on the health-related quality of life in patients with type 2 diabetes and comorbidity. Diabetes Care 2009;32(9):1594-6.

22. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the Chronic Care Model in the new millennium. Health Aff 2009;28(1):75-85.
23. Saaddine JB, Cadwell B, Gregg EW, Engelgau MM, Vinicor F, et al. Improvements in diabetes processes of care and intermediate outcomes: United States, 1988 - 2002. Ann Intern Med 2006;144(7):465-74.
24. Hedblad B, Nerbrand C, Ekesbo R, Johansson L, Midlöv P, et al. High blood pressure despite treatment: results from a cross-sectional primary healthcare-based study in southern Sweden. Scand J Prim Health Care 2006;24(4): 224-30.
25. Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic health records and quality of diabetes care. N Engl J Med 2011;365(9):825-33.
26. van Lieshout J, Frigola Capell E, Ludt S, Grol R, Wensing M. What components of chronic care organisation relate to better primary care for coronary heart disease patients? An observational study. BMJ Open 2012;2(4):e001344.
27. Russell GM, Dahrouge S, Hogg W, Geneau R, Muldoon L, et al. Managing chronic disease in Ontario primary care: the impact of organizational factors. Ann Fam Med 2009;7(4):309-18.
28. Adolfsson ET, Smide B, Rosenblad A, Wikblad K. Does patient education facilitate diabetic patients' possibilities to reach national treatment targets? Scand J Prim Health Care 2009; 27(2):91-6.

APPENDIX

Appendix 1. Creator of composite scores

Twelve organizational determinants were identified in this study. To create a composite measure out of these 12 determinants we transformed six measures into dichotomous measures. These determinants were nurse practitioner volume, patient awareness of the complaints procedure, use of the electronic medical records, consultation with partners, collaboration with partners, and use of patient leaflets. We based the division into two groups on the median score within our dataset. Since in several cases the median score equaled either the highest or lowest score, we looked at the average to determine in which group the median score should be included. If the average was lower than the median, we coded all median or higher scores as 1 and the others as 0, and vice versa. For example, 61% of all practices scored 0 on the determinant consultation with partners, 24% scored 1 and 11% scored 2. We recoded this into '0' for all practices with a 0 score, and '1' for all other practices. The other six measures were already dichotomous and coded as 0 or 1. All 12 items were then combined into a sum score that indicated the extent of implementation of all determinants. This score could in theory vary between 0 (none of the determinants implemented) and 12 (all determinants implemented).

Appendix 2. Multilevel models (cluster measure = practice, data on patient level) on continuous outcomes of HbA1c, cholesterol and blood pressure for separate determinants

Parameter*	HbA1c (%), n = 7281		Total cholesterol (mmol/l), n = 7122		Systolic blood pressure (mm Hg), n = 7320	
	Estimate (95% CI)	*p*	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*
Intercept	7.28 (6.84 – 7.72)	.000	4.54 (4.08 – 5.00)	.000	119.8 (112.3 –127.3)	.000
Volume of NP◊	.083 (-.518 – .684)	.786	-.088 (-.712 – .535)	.780	-5.30 (-15.39 – 4.78)	.301
Annual report	.103 (.009 – .197)	.032	.019 (-.078 – .117)	.703	-.23 (-1.81 – 1.35)	.774
Complaints policy	-.002 (-.005 – .002)	.409	.001 (-.003 – .005)	.614	.02 (-.04 –.08)	.521
Use of EMR◊	-.088 (-.130 – -.045)	.000	.007 (-.037 – .051)	.759	-.56 (-1.27 – .15)	.124
Diabetic clinic	.327 (.173 – .480)	.000	.056 (-.103 – .214)	.489	.81 (-1.77 – 3.38)	.538
Consultation	-.013 (-.048 – .023)	.482	-.003 (-.040 – .034)	.872	-.41 (-1.00 – .19)	.181
Collaboration	-.042 (-.109 – .025)	.215	-.006 (-.075 – .063)	.867	-.13 (-1.26 – .99)	.816
FP education	.020 (-.063 – .103)	.639	-.001 (-.087 – .086)	.991	-.51 (-1.91 – .88)	.470
Nurse education	-.027 (-.115 – .061)	.542	-.034 (-.125 – .057)	.464	-.79 (-2.27 – .68)	.292
Guidelines	.152 (-.093 – .397)	.222	-.008 (-.261 – .245)	.949	.34 (-3.77 – 4.45)	.870
Patient leaflets	.053 (-.031 – .138)	.215	-.012 (-.100 – .075)	.784	-1.07 (-2.49 – .345)	.137
Reminder system	-.005 (-.113 – .103)	.928	.047 (-.065 – .158)	.410	-1.35 (-3.16 – .-46)	.144
Patient age	-.002 (-.004 – -.000)	.034	-.001 (-.003 – .001)	.431	.34 (.31 – .37)	.000
Patient gender	.028 (-.018 – .073)	.230	-.353 (-.402 – -.305)	.000	-.53 (-1.34 – .27)	.195
Year data collection						
2007	.056 (-.056 – .168)	.325	-.001 (-.117 – .114))	.986	2.69 (.80 – 4.58)	.005
2008	.062 (-.056 – .180)	.299	-.155 (-.277 – -.032	.014	2.53 (.54 – 4.51)	.013
2009	.129 (-.046 – .303)	.148	-.072 (-.252 – .108)	.431	2.15 (-.78 – 5.08)	.149
Practice type						
Single handed	.045 (-.072 – .163)	.447	-.042 (-.164 – .080)	.495	-1.51 (-3.49 – .46)	.132
Health center	.115 (-.010 – .241)	.070	.133 (.003 – .263)	.045	.00 (-2.10 – 2.10)	1.00

** Reference categories dichotomous measures: amount of nurse education below median, amount of family physician (FP) education below median, no diabetic clinic, practice does not produce an annual report, no reminder system available, patient gender male, data collection in 2006, practices with more than 1 FP that are not part of a health center. ◊ NP = Nurse practitioner, EMR = electronic medical records.*

8

General discussion

'Measurement is the first step that leads
to control and eventually to improvement.
If you can't measure something, you can't understand it.
If you can't understand it, you can't control it.
If you can't control it, you can't improve it.'

H. James Harrington

BACKGROUND

This thesis focuses on the improvement of chronic care management in primary care, using practice accreditation for audit and feedback. Audit and feedback gives an opportunity to reflect on clinical and organizational performance by comparing current performance with best practices or prior performance. Despite several decades of research, the heterogeneity of effects of this strategy remains large and not well understood[1]. Practice accreditation is a specific type of audit and feedback, in which results are used to grant accreditation to the practices[2]. The research on the effectiveness is hindered by the major differences in the content of the various programs[3]. In this thesis, one component of the Dutch Accreditation program for general practices is studied, namely the indicators and feedback concerning clinical processes and outcomes of general practice care for patients with chronic diseases. Little was known about the metric properties of these clinical quality indicators based on national guidelines. We therefore examined the reliability and validity of the audit instruments, as well as the relation between sample size and precision. Subsequently, the clinical indicators were used to analyze the impact of the Dutch Accreditation program on chronic care management. This study was complemented by a qualitative study to learn more about the factors and processes related to change. Another issue that needed to be addressed was the influence of the specific characteristics of each practice population on the impact of the Accreditation program. In the calculation of clinical performance scores, it has been suggested to account for patient characteristics that can be distributed differently between practices[4,5]. As the practical relevance of risk adjustment for patient characteristics is unknown, we investigated its impact on the scores. To conclude the thesis, an in-depth analysis was performed on the organizational determinants that affect chronic care management in primary care. Most of these factors have been studied in controlled trials[6], but it was questioned whether they had impact in routine diabetes care as well.

MAIN FINDINGS

Reliability and validity of the measurement instruments

- ✓ The clinical quality indicators used in the Dutch Accreditation program measure a number of conditions and topics that show low to moderate correlations (Pearson correlations up to 0.48), suggesting that they measure a broad scope of provided care. Indicators on a single condition show reasonable internal consistency (Cronbach's alpha between 0.56 and 0.73), which suggests that the indicators measure one underlying concept. These results indicate that the instruments used in the Accreditation program are valid for their purposes: formative evaluation and self-directed learning.

- ✓ Given the sample of 40 patients used in the Dutch Accreditation program, a confidence limit of 15% needs to be taken into account around performance scores. For example, the underlying true score of a practice could be anywhere between 52% and 82% when the measured score is 67%. Current confidence limits around benchmarks are smaller, around 5%. These confidence limits seem to be acceptable for the purpose of internal quality performance. However, in case of summative assessment or pay-for-performance, higher levels of precision are required and sample sizes should be increased. Given the measurement goal, a balance should be found between an acceptable level of precision and the increased burden of data collection. The aggregation level (practice or group of practices) is also relevant in this context.

Precision of the measurement instruments

- ✓ The number of cases required to achieve a certain level of precision can be reduced by creating a composite score that is based on several single indicator performance scores. Whereas for a single indicator a sample size of 96 cases is required to achieve a 10% precision level, this sample size can be reduced to between 43 and 89 cases for a composite performance score. The size of the reduction is dependent on the strength of the correlations between the indicators in the composite and on their scores. But even for composite scores, the sampling of large numbers of patient records would be required to achieve high levels of precision, that is a confidence limit of 5% or less.
- ✓ Collecting data on only a small number of indicators per condition does not necessarily lead to lower precision. A composite score based on an optimally chosen subset of 5 or 6 indicators can have the same level of precision as a composite score that is based on the full set of 9 to 12 indicators.

Impact of participation in the Accreditation program

- ✓ Participants in the Dutch Accreditation program that were measured with 24 indicators at start and completion of a three-year period had improved on 6 process indicators related to diabetes (feet examination, cholesterol measurement, lipid lowering medication prescription) and COPD (spirometry measurement, smoking cessation advice). Compared to practices that had just started in the Accreditation program, they also performed better on 4 indicators related to diabetes (cholesterol outcome) and cardiovascular disease (blood pressure outcome, smoke status registration, glucose measurement).
- ✓ However, the improvements could not be directly related to participation in the Dutch Accreditation program, possibly due to the effects of other quality improvement initiatives that were applied simultaneously.

Experiences of participants in the Accreditation program

- ✓ Family physicians primarily chose to participate in the Dutch Accreditation program because they wanted to improve the quality of care. They were positive about the insight they gained into the clinical care they provided and their practice management. The main drawback of the process mentioned by the physicians was the amount of time and effort it took to collect data. Furthermore, they suggested that they needed more support in preparing a practice improvement plan.
- ✓ Some physicians considered the drawbacks of the process to outweigh its advantages. In order to optimize quality of care measurement, the findings suggest that it is crucial to find a balance between measurability and perceived relevance.

Patient characteristics associated with performance scores

- ✓ Several patient characteristics were associated with indicators of diabetes care. Moderate effects of patient characteristics were found on processes of diabetes care (measurement of HbA1c, blood pressure and LDL cholesterol), with areas under the curve that varied between 0.66 and 0.76. The main contributor to these effects was the frequency of diabetes related consults, that can be seen as a measure of patient compliance to the treatment process.
- ✓ Regarding the intermediate outcomes of diabetes care (HbA1c, blood pressure and LDL cholesterol values within target levels) we found considerably smaller effect sizes, with areas under the curve between 0.57 and 0.61. These results suggest that correction or stratification of patient characteristics may have limited practical relevance.

Organizational determinants associated with performance scores

- ✓ Most investigated organizational determinants were broadly implemented in routine care, but only a few (3 out of 36) showed significant effects on intermediate diabetes outcomes. Associations were found between the use of electronic medical records and HbA1c, the availability of patient leaflets and blood pressure control and the number of hours of nurse education and cholesterol control.
- ✓ A higher score on the composite measure of several practice determinants was associated with better systolic blood pressure control. These results from an observational study of routine primary care support the findings of trials on organizational changes in diabetes care.

STRENGTHS AND LIMITATIONS

The research in this thesis was mainly based on the data collected in the Dutch Accreditation program. These data specifically lend themselves for observational studies. Whereas randomized, controlled trials (RCTs) have obvious advantages over observational studies in terms of risk of bias, observational studies can be used to complement evidence from trials to enhance the generalizability. Results from observational studies in daily practice can be more easily translated to this setting. Although some studies show that effect sizes in RCTs tend to be smaller than in observational studies[7], Anglemyer et al. found little evidence for differences in effect sizes in their systematic review[8]. Nevertheless, there were several limitations in our studies that relate to the use of observational data. During the data collection period, changes were made to the instruments as a reflection of changes in guidelines and policy. This led to issues with comparability of the data, especially in our study of impacts of practice accreditation. Furthermore, it is likely that practices with a less than average interest in quality improvement may have been underrepresented in our sample, although practice characteristics of participating practices were comparable to all Dutch practices. Some of the practices participated in a pay-for-performance program, which was run parallel in time; Kirschner et al. showed that these practices had improved on both process and outcome measures of chronic care management[9]. However, in our study of impacts, we did not find any differences between the pay-for-performance participants and the other practices that took part in the Dutch Accreditation program. For the study on the experiences of the practices, data collection took place in collaboration with the project on pay-for-performance, meaning that only practices that participated in both studies were included. This may have biased the results, although it is unclear in which direction.

Our dataset was extracted retrospectively from the electronic medical records. In most studies, we addressed the issue of missing values by selecting a subsample of completed data. This extra sampling process was checked on loss of validity by comparing the patient and practice characteristics of the selected samples with the total dataset. However, in the study on the impact of patient characteristics we used multiple imputation to account for the missing data due to the large amount of missing data on some of the variables.

In our study population, the overall quality of chronic care management in the practices that participated was already reasonably high, especially regarding diabetes. This may have been caused by previous improvement initiatives or by the voluntariness of the program. Possibly, the high levels of quality may have hindered finding significant effects in our study on impacts of accreditation and the studies on the association of patient and practice characteristics, because it becomes more difficult to establish significant effects when baseline scores are relatively high[6].

PERFORMANCE MEASURES

Feedback in the Accreditation program is provided through performance scores calculated from clinical quality indicators. One of the goals of this thesis was to gain better insight in the reliability and validity of the instruments. In this section, we address several aspects that are of importance in the evaluation of the measurement instruments and what conclusions can be drawn from performance scores. We found that electronic record data are sufficiently reliable to measure quality of chronic care management for internal assessment. This is supported by Green et al.[10], who found that information from electronic medical records is especially useful to evaluate performance measures of technical quality of care such as chronic care management. Nevertheless, Parsons et al.[11] showed that information derived automatically from the electronic medical records can be biased towards lower performance scores. Measured effects may be a reflection of the quality of registration rather than the quality of care. In the Dutch Accreditation program, physicians often extracted data from their medical records manually. This meant that information that was recorded incorrectly, i.e. in other record fields than it was expected to be recorded in, could still be retrieved and included. Therefore, in our study manual extraction has probably decreased the bias described by Parsons. On the other hand, this option to manually extract data also meant that smaller samples of patients were included per practice. As we have shown in our study on the precision of data, smaller samples lead to an increase in the confidence limit around the scores. In other words, the sample may be a less adequate reflection of the practice population because of larger random error. This is important information that can be used to improve the interpretability of the performance scores.

Regarding the evaluation of the measurement instruments, physicians were instructed in the data collection protocol to leave out patients that had a medical specialist as their main care provider. To judge chronic care management fairly, it makes sense to only include those patients for which the family physician (and nurse practitioner) make the main decisions regarding treatment. However, this may have caused bias, systematic error in the reflection of the population, because the patients that are treated by a specialist tend to have a more severe or less stable condition[12]. Although a written protocol was used to insure random selection of patients, there may also have been a bias in patient selection with a preference for well-controlled patients. However, in comparing our performance scores on diabetes to data from the Dutch GIANTT study[13] – a study in which all diabetes patients were included that have the FP as their main diabetes care provider – similar results were shown. This suggests that in our study, selected patient samples were an adequate reflection of the practice population.

Another point of discussion relates to our studies on the impact of various patient characteristics on processes and outcomes of care. In this thesis, we found that the impact of these factors was small to moderate. Other studies have shown that effects can be small for most but large for a few practices[5,14]. Correction or stratification of performance scores can increase their face validity and thereby the acceptance among professionals and stakeholders[15]. However, it is debatable whether, for an individual practice, the benefit of correction outweighs the disadvantages of case-mix, which include complex calculation methods and more difficult interpretation of scores. Whichever choice is made, it is important to provide adequate information for users to interpret performance scores correctly[16].

We used classical (frequentistic) statistics to analyze our data in this thesis. In addition, one could also use Bayesian methods. A Bayesian analysis is based on several probability statements: a prior distribution that is subjective, and a posterior distribution that is based on the prior distribution and the observed data[17]. The probability of whether the blood pressure of a diabetes patient is within target level in a certain practice can be estimated, for example based on guidelines. These probabilities are then adjusted with the incorporation of the observed data, after which the fit of the model can be assessed. One advantage of Bayesian methods is that it allows for expert or stakeholder judgments in an explicit way; conclusions of an analysis may vary depending on the context of the study and who is conducting it[18]. This may be beneficial, for example when you want to incorporate the different viewpoints of various stakeholders. Several separate prior distributions can be included that correspond with each stakeholder. Another advantage of Bayesian models is that small sample sizes (per family physician or per practice) do not have to pose a problem, since the model is fit by pooling data across all physicians or practices[19].

Implications

This thesis includes a number of studies that aim to enhance our knowledge on the use of performance scores and their interpretation in practice. Several implications can be derived for practical use of performance scores. The current performance scores are based on relatively small sample sizes; an acceptable level of error needs to be taken into account. Although precision can be increased by the creation of a composite score, the extent of this increase is highly variable between practices. Precision can also be increased by collecting data from larger samples; improved registration and facilitation of data extraction can contribute to achieve larger samples from individual practices. Or data can be collected on a larger scale, for example on care trust level. However, for an individual practice, the key is to accept a level of uncertainty around each performance score. Knowledge on the issues raised in this thesis can help to draw correct conclusions from the performance scores. While insight in the impact of case-mix factors and organizational determinants is important, based on this

insight one may choose not to incorporate these factors into practice. Instead, uncorrected or non-stratified scores may be acceptable for internal quality improvement since the impact of these factors is limited. Further studies on validity and reliability of performance scores, based on frequentist or Bayesian statistics, can be beneficial to gain more knowledge. Statistics can be an important tool to support decisions on which quality indicators to include and what conclusions can be drawn from performance scores. However, the means should not exceed the goal. It is crucial to align statistical issues of assessment to issues of clinical relevance to professionals, patients and policy makers, to increase the chances to influence professional behavior[20]. Bayesian statistics may offer a solution by combining the input from experts and statistical models, and thereby it may provide interesting additional information to evaluate performance measures.

AUDIT AND FEEDBACK IN AN ACCREDITATION PROGRAM

The Dutch Accreditation program uses audit and feedback as a central mechanism, embedded in a continuous quality improvement process[21]. This process aims to enhance a culture that fosters quality and safety of healthcare[22]. Literature shows that audit and feedback can have positive effects on processes and outcomes of care, although effect sizes are heterogeneous and overall small to moderate[23]. In this thesis, we found several improvements on processes and outcomes of care, but the association with the Accreditation program was not obvious and effect sizes were small. We also found mixed results in our qualitative study. Therefore we did not find support for an added value of the Dutch Accreditation program to the effects of audit and feedback, as compared to usual care and other ongoing programs, that could be expected on the basis of some previous research[23]. Our findings are in line with systematic reviews on the effects of accreditation on quality of chronic care[21]. However, there may be more indirect effects of the Accreditation program that we have not measured with our instruments and within the timeframe of the study. Nouwens et al.[24] also investigated the effects of the Dutch Accreditation program; they specifically looked at the added value of the improvement plans on cardiovascular risk management. They found no effects on the outcomes, but found improvements on six processes of care: recorded smoking status, exercise control, diet control, registration of alcohol intake, measurement of waist circumference and fasting glucose. In an additional qualitative study, Nouwens et al. found that physicians did not perceive any direct effects of the Accreditation program on patient care. However, they did report positive effects on team climate and on the commitment to improve quality of care[25].

A systematic review on audit and feedback states that effects can be larger when feedback is provided more intensively, when providers are actively involved and have specific and formal responsibilities for implementing change[23]. One of the key features of the Dutch Accreditation

program is that the practice is accredited if it is actively involved in quality improvement through use of improvement plans and can achieve improvements over time. This approach should aid in creating active involvement among the health professionals; they can take ownership of the improvement plans that can be tailored to the individual practices. However, results from our qualitative effect study do not support that this was the case. Although physicians felt positive about the opportunity to individualize the improvement plans to their practice, they did not take ownership of the improvement plans; one of the reasons being that they felt that the methods used to develop the plans did not suit their pragmatic way of thinking. Possibly, other barriers on the level of the healthcare provider, team or organization (such as a less than optimal team climate or lack of leadership) may have hindered the implementation of practice changes as well[26]. In their qualitative study[25], Nouwens et al. investigated which determinants were perceived to facilitate implementation of the Dutch Accreditation program. These include a heightened enthusiasm for improving quality of care, clear communication and designation of one person to take responsibility for the program.

Accreditation programs vary widely in their content. The Dutch Accreditation program has been developed for internal quality assessment. However, there is an increasing request to disclose quality data for the general public. Enhancing transparency has been a key element in changes in various healthcare systems in the past decade[27] and remains high on the agenda in the current decade. Transparency can be used to prove that care is provided according to the standards and may help to give confidence that quality is high in certain areas. Furthermore, it is thought to have an effect on improvement through the mechanism that patients choose for higher scoring practices[28]. A systematic review[29] that is mainly based on studies performed in hospitals shows mixed results on publishing patient care data. Given the issues on validity and precision of performance scores mentioned in the section on performance scores, we suggest that additional testing of the indicators is necessary before data can be used for public disclosure and external comparisons. Furthermore, it is questionable whether the benefits of transparency outweigh the costs and effort to collect data when the direct association with improved processes and outcomes of care is not clear.
Information can also be used in a pay-for-performance scheme. Kirschner et al.[30] have looked at effects of the Dutch Accreditation program on chronic care, but with an additional financial incentive provided based on performance. They found that practices had improved their care within the time period measured. It is not certain whether effects can be sustained over a longer time period[9]. However, there can also be negative consequences to financial incentives. The implementation of the Quality Outcomes Framework (QOF) in the UK provides a good example of what can happen when there is a strong focus on outcomes in

combination with relatively large incentives. Researchers found that there were several unintended consequences, including gaming, a lack of improvement on non-incentivized indicators and a slight decrease in scores after incentives stopped. Other aspects such as accessibility also worsened in some cases[31,32]. They conclude that part of these unintended consequences were caused by the large incentives that were involved. These issues will be more prominent in external summative assessment compared to internal quality improvement.

Implications

Since we performed this study, changes have been made to facilitate the process and make it easier to tailor the programs to the practice, which may have helped to broaden the support among physicians. The focus of the program has also shifted: whereas the aim used to be to achieve the highest indicator scores possible, currently practice scores are compared to certain benchmark levels that indicate a minimum level of good quality of care. In the future, the program would benefit from further facilitation of data collection. Also, chronic care management has been divided into separate modules that can be spread over several years. This decreases the amount of data collection in one year and offers the physicians the option to focus on one or a few topics at a time. Further efforts should be made to enhance the chances for the Accreditation program to be effective, including taking into account barriers and facilitators in improvement that have not been taken into account in the current program, for example team climate and motivation to change[26]. Furthermore, it is important to work towards one measurement instrument that different stakeholders can use to gain insight in performance. Currently, in the Netherlands several different indicator sets are used in care trusts, the Accreditation program and other initiatives to accumulate data for stakeholders such as the health inspection and insurers. The practices would benefit from one indicator set that is sufficient to provide data for all these different purposes. A standard that aims to promote this is under development at the Netherlands Care Organization (Zorginstituut Nederland).

Guidelines, from which performance scores are derived, have increasingly become a golden standard. Each deviation needs to be accounted for by the professional. It is important to be aware of the original and primary goal of a practice guideline: to guide the physician, rather than to enforce strict rules. As a professional, the physician is responsible to provide the best, tailored, treatment for an individual patient. They can take an active role in quality improvement; they are also the most suitable person to decide in individual cases whether to deviate consciously from a guideline or not. Indicators on the percentage of patients treated according to guidelines are meant to indicate and should not be used in black-and-white

comparisons. Instead of the strong focus on the actual outcomes, the Dutch Accreditation program was designed to facilitate systematic quality improvement in practices.

CHRONIC CARE MANAGEMENT

This thesis focuses on chronic care management. Performance of primary care practices is measured with indicators that are based on national guidelines[33-35]. In our study on the impact of accreditation, processes of care were in accordance with guidelines in approximately 50-90% of the patients (highest scores on diabetes indicators and lower scores on chronic obstructive pulmonary disease (COPD) and cardiovascular disease (CVD) indicators). On the other hand, guideline recommendations were met in 35-65% of the outcome measures. These findings are in line with other national studies[13,36]. These percentages show that there is room for improvement, especially on the outcomes. However, the exact improvement ratio is unclear: for most indicators it is not desirable to strive for 100% scores, i.e. treatment according to the guideline in each patient. There are several reasons for lower guideline adherence. First of all, the guidelines themselves vary in the amount of evidence and the urgency of the recommendations. For example, the guidelines state that periodic control of diabetes patients that are not fully under control should occur (at least) 4 times per year[33]. However, scientific evidence for this recommendation is lacking. In this thesis, the study on impact of patient characteristics shows that higher visiting frequencies are associated with better processes of care, which would suggest that one should increase the number of consultations when patients are not under control properly.

Another reason for lower guideline adherence is that throughout the years several changes have been made to the guidelines. For instance, target levels have changed for diabetes and CVD outcomes. Other changes have also been implemented as the evidence changed. For example, whereas previously spirometry was suggested as a tool for both diagnosis and follow up, it is now suggested mainly as a diagnostic tool and to measure large changes in lung function[37]. Treatment itself is guided mostly by patient complaints rather than clinical parameters. Given these changes in guidelines, indicators are adjusted as well. Although this offers some challenges in the interpretation and comparability of the collected data, performance scores should reflect the guidelines as closely as possible in order to achieve the most optimal reflection of good quality of care given the evidence at the time of measurement. However, it takes time to change the guidelines and performance measures according to the available evidence, so that at any moment in time, the most recent information may not be taken into account yet.

A systematic review[38] reports that poor guideline adherence can also be associated with barriers on patient and physician level, including their attitudes, beliefs and their knowledge. In individual cases, guideline adherence may be lower because physicians may choose to

deviate consciously from the guideline for clinical reasons. For example because further intensification of medical treatment would make the situation worse, despite the fact that treatment targets were not achieved. On the other hand, physicians may also deviate from guidelines to achieve better patient compliance (e.g. patients are more easily motivated to aim for a weight loss of 5 kilos than 30 kilos). Treatment can also be guided by the severity of the symptoms and preferences of the patients, which can also lead to deviations from the guidelines.
The interpretation of performance scores on chronic care management can be challenging. Literature shows that several patient characteristics are associated with processes and outcomes of care[39,40]. In this thesis, we found small associations of several patient characteristics with performance scores. Given these influences we suggest that indicator scores may be adjusted for these factors through stratification rather than correction. The benefit of stratified scores is that physicians gain additional information on whether they deviate from the guidelines in a particular group of patients. In addition, continued explanation of the limitations of performance and outcome data to relevant stakeholders may be required.

The performance measures on chronic care management are based on the guidelines and validated with expert panels[41]. However, there are other aspects of quality that we have not measured with our set of performance scores. Indicators only form a snapshot of the care at one moment in time, whereas good quality of care also encompasses timeliness[42]. Sidorenkov et al.[43] investigated the differences in guideline adherence for commonly used indicators, such as those in this thesis, versus clinical pathways in diabetes care. They found that the quality of care processes was substantially reduced when looking at clinical pathways compared to process indicator scores. For example, in their study 86% of all diabetes patients received a blood pressure measurement. However, they found that only 53% of all diabetes patients were adequately managed in response to this measurement (i.e. received treatment intensification when indicated and had a timely evaluation of their response to treatment). Porter takes the idea of clinical pathways even further and argues that accountability should be shared among the providers involved[44]. The focus should be on all care for a medical condition, including its complications; involving hospital an general practice, but also other care providers such as physiotherapists and district nurses. Furthermore, he provides a hierarchical model of outcomes with three tiers: retained or achieved health status, process of recovery and sustainability of health. Each condition may require different outcomes on the three tiers. The relevance of each tier may also vary: for example, in advanced cancer stages providers may have limited effects on survival and other outcomes related to retained or achieved health status[44]. Whichever method is used to

measure quality of care, it will always be a simplified version – an indication – of reality. This offers challenges in whether measures are an adequate reflection of care. In our interview study, some physicians commented that the instruments of the Accreditation program on practice organization did not capture actual quality of care. In order to increase the support among users, the instruments in the accreditation are regularly updated, in accordance with the latest evidence.

Another aspect of chronic care that has not been measured is the actual functioning of the patient in daily life. Although patient-reported outcome measures (PROMs) are most often used when a clear comparison can be made, for example before and after a surgical procedure, they can also be beneficial for quality measurement of chronic care[45]. For example, a self-reported quality of life questionnaire may offer insight in the patient's fears of going into a hypoglyceamic coma and its effect on everyday life[46].

In recent years, personalized medicine has been proposed as an important future perspective of healthcare[47]. Recently developed guidelines include variable targets, for example with a HbA1c target level in diabetes patients that varies according to age, duration of the disease and treatment intensity, as opposed to a fixed target that is valid for all patients[33]. However, others suggest that guidelines should be adjusted further, by incorporating patient preferences such as their beliefs, expectations and goals[48]. This causes some issues for mechanisms such as accreditation, which strives for standardization of treatment, a minimum standard or an optimal treatment to strive for in each patient. In some cases a higher diversity in treatment, resulting in higher practice variation, is actually an indication of good quality of care. With the current quality indicators, we are not able to distinguish between intentional deviation in order to benefit the patient and unintentional deviation (i.e. lack of good quality care). However, they can still offer evaluative information that can form the basis to discuss the processes that take place within the practice.

Implications

Given the fact that the translation of performance measures into improvements in chronic care management is not straightforward, there are several implications for practice. Striving for 100% scores may not always be in the patient's best interest, so indicator scores should be assessed with a view on this. It is also important to continuously develop the guidelines and measurement instruments according to the latest evidence and to allow for conscious deviation from the guidelines in individual cases. For physicians, the challenge is to be prepared for these changes that will affect patient management. This may be particularly challenging if certain procedures were previously implemented in the guideline, but are later on discarded. For this was the case with spirometry as a measure to guide treatment of COPD. It is sensible to use performance measurements purely as an indication.

The increased administrative burden that comes with accreditation and other improvement initiatives that require data collection may have a negative effect on patient experiences with the consultation. Shachak et al.[49] found that patient-doctor communication was hindered by the use of a computer to fill in electronic medical records during consultations. The benefits of the program should be weighed against these issues. The aim should be that physicians can collect data that they would normally collect to provide good quality of care, with a minimum extra impact from improvement schemes. Extra efforts are put in to ensure better possibilities for automatic data extraction in the future.

With the increasing focus on personalized medicine, deviations from the guidelines that are well thought out may become more frequent, thus potentially increasing practice variation. In the development of new indicators this should be taken into account. As an addition to the existing indicator sets on diabetes, COPD, CVD and asthma, we could develop indicators for other common diseases such as depression. Furthermore, patient reported outcome measures (PROMs) may offer insight in the impact of the chronic disease on everyday life. Also, in line with patient-centered care, indicators could be developed regarding multi-morbidity or collaborative care.

CONCLUSIONS

Although the effects of accreditation found in this thesis are limited, an accreditation program remains a suitable tool for quality improvement. Our research on the properties of the indicators and effects of patient and practice characteristics provides evidence on the impact of a specific accreditation program. We should strive for the optimal balance between the creation of accurate and reliable performance measures versus acceptable amounts of investments to collect data. Formative usage of the Accreditation program can be recommended; after further evaluation, the program as a whole could be suitable for a summative evaluation, in which for example the impact of improvement actions are assessed. Guidelines and indicators are aimed to do nothing more or less than to guide and to indicate, so technical knowledge as well as wisdom is required in the use for quality improvement.

REFERENCES

1. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, et al. Audit and feedback: effects on professional practice and health-care outcomes. Cochrane Database Syst Rev 2012;6:CD000259.
2. Shaw C. The external assessment of health services. World Hosp Health Serv 2004;40(1): 24-7, 50, 51.
3. Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review. Int J Qual Health Care 2008;20(3):172-83.
4. Paddison C, Elliott M, Parker R, Staetsky L, Lyratzopoulos G, et al. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. BMJ Qual Saf 2012;21(8):634-40.
5. Zaslavsky AM, Hochheimer JN, Schneider EC, Cleary PD, Seidman JJ, et al. Impact of sociodemographic case mix on the HEDIS measures of health plan quality. Med Care 2000;38(10):981-92.
6. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet 2012;379(9833):2252-61.
7. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286(7):821-30.
8. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev 2014;4:MR000034.
9. Kirschner K, Braspenning J, Akkermans R, Jacobs J, Grol R. Assessment of a pay-for-performance program in primary care designed by target users. Fam Pract 2013;30(2):161-71.
10. Green ME, Hogg W, Savage C, Johnston S, Russell G, et al. Assessing methods for measurement of clinical outcomes and quality of care in primary care practices. BMC Health Serv Res 2012;12:214.
11. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. J Am Med Inform Assoc 2012;19(4):604-9.
12. Sluiter AC, van Wijland JJ, Arntzenius AB, Bots AF, Dijkhorst-Oei LT, et al. Landelijke transmurale afspraak diabetes mellitus type 2. Huisarts Wet 2012;55(1):S1-12.
13. Voorham J, Haaijer-Ruskamp FM, van der Meer K, de Zeeuw D, Wolffenbuttel BH, et al. Identifying targets to improve treatment in type 2 diabetes; the Groningen initiative to analyse type 2 diabetes treatment (GIANTT) observational study. Pharmacoepidemiol Drug Saf 2010;19(10):1078-86.
14. Hong CS, Atlas SJ, Chang Y, Subramanian SV, Ashburner JM, et al. Relationship between patient panel characteristics and primary care physician clinical performance rankings. JAMA 2010;304(10):1107-13.
15. Eselius LL, Cleary PD, Zaslavsky AM, Huskamp HA, Busch SH. Case-mix adjustment of consumer reports about managed behavioral health care and health plans. Health Serv Res 2008;43(6):2014-32.
16. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest 2007;8(2):53-96.
17. Bolstad WM. Introduction to Bayesian statistics. Hoboken, NJ: John Wiley; 2007 (2nd ed.).
18. Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. Statist Sci 2004;19(1):156-74.
19. Austin PC. A comparison of Bayesian methods for profiling hospital performance. Med Decis Making 2002;22(2):163-72.
20. McDonald R, Roland M. Pay for performance in primary care in England and California: comparison of unintended consequences. Ann Fam Med 2009;7(2):121-7.
21. Alkhenizan A, Shaw C. Impact of accreditation on the quality of healthcare services: a systematic review of the literature. Ann Saudi Med 2011;31(4):407-16.
22. Greenfield D, Pawsey M, Braithwaite J. What motivates professionals to engage in the accreditation of healthcare organizations? Int J Qual Health Care 2011;23(1):8-14.
23. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback:

effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2006;2:CD000259.
24. Nouwens E, van Lieshout J, Bouma M, Braspenning J, Wensing M. Effectiveness of improvement plans in primary care practice accreditation: a clustered randomized trial. Plos One 2014;9(12):e114045.
25. Nouwens E, Van Lieshout J, Wensing M. Determinants of change in a practice accreditation program in primary care: a qualitative study. Submitted.
26. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003;362 (9391):1225-30.
27. Christianson JB, Volmar KM, Alexander J, Scanlon DP. A report card on provider report cards: current status of the health care transparency movement. J Gen Intern Med 2010;25(11):1235-41.
28. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. JAMA 2005;293(10):1239-44.
29. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008; 148(2):111-23.
30. Kirschner K, Braspenning J, Jacobs J, Grol R. Design choices made by target users for a pay-for-performance program in primary care: an action research approach. BMC Fam Pract 2012;13:25.
31. Roland M. QOF – too much of a good thing? Cambridge: Cambridge Centre for Health Services Research, 2013.
32. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ 2011;342:d3590.
33. NHG-standaard diabetes mellitus type 2 (derde herziening). Huisarts Wet 2013;56(10): 512-25.
34. NHG-standaard COPD (tweede herziening). Huisarts Wet 2007;50(8):362-79.
35. NHG-standaard cardiovasculair risicomanagement (eerste herziening). Huisarts Wet 2012; 55(1):14-28.
36. Landelijke Organisatie voor Ketenzorg. Transparante ketenzorg: diabetes mellitus, COPD en VRM. Rapportage zorggroepen over 2012. Utrecht: LHV/VHN/LVG; 2013.
37. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Pocket guide to COPD diagnosis, management, and prevention: a guide for health care professionals. GOLD; 2014.
38. Nam S, Chesla C, Stotts NA, Kroon L, Janson SL. Barriers to diabetes management: patient and provider factors. Diabetes Res Clin Pract 2011;93(1):1-9.
39. Malik S, Lopez V, Chen R, Wu W, Wong ND. Undertreatment of cardiovascular risk factors among persons with diabetes in the United States. Diabetes Res Clin Pract 2007;77(1):126-33.
40. Schaars CF, Denig P, Kasje WN, Stewart RE, Wolffenbuttel BH, et al. Physician, organizational, and patient factors associated with suboptimal blood pressure management in type 2 diabetic patients in primary care. Diabetes care 2004;27(1):123-8.
41. Braspenning J, Dijkstra R, Tacken M, Bouma M, Witmer H. Visitatie Instrument Accreditering (VIA®). Nijmegen/Utrecht: Kwaliteit van Zorg (WOK); UMC St Radboud; Nederlands Huisartsen Genootschap (NHG); NHG Praktijk Accreditering BV (NPA); 2007.
42. Committee on Quality of Health Care in America; Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.
43. Sidorenkov G, Haaijer-Ruskamp FM, de Zeeuw D, Denig P. A longitudinal study examining adherence to guidelines in diabetes care according to different definitions of adequacy and timeliness. Plos One 2011;6(9):e24278.
44. Porter ME. What is value in health care? N Engl J Med 2010;363(26):2477-81.
45. Santana MJ, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. Qual Life Res 2014;23(5):1505-13.
46. Meadows KA. Patient-reported outcome measures: an overview. Br J Community Nurs 2011;16(3):146-51.
47. Capko J. The patient-centered movement. J Med Pract Manage 2014;29(4):238-42.

48. Montori VM, Brito JP, Murad MH. The optimal practice of evidence-based medicine: incorporating patient preferences in practice guidelines. JAMA 2013;310(23):2503-4.

49. Shachak A, Reis S. The impact of electronic medical records on patient-doctor communication during consultation: a narrative literature review. J Eval Clin Pract 2009;15(4):641-9.

Summary

'Quality means doing it right when no one is looking.'

Henry Ford

This thesis examines practice accreditation as a strategy to improve chronic care management in primary care. Practice accreditation is offered in many countries, but little is known on its effectiveness. Evaluation of chronic care management is often an important part of these accreditation programs. In this thesis the measurements themselves are discussed as well as the impact of the Dutch Accreditation program on the performance improvement.

Chapter 1 provides a general introduction of chronic care management and practice accreditation programs. Chronic care management increasingly becomes a more important part of the care provided by general practices. It is predicted that the prevalence of chronic conditions, as well as the number of people with multiple chronic conditions, will increase in the coming years mainly due to an aging population. This will lead to a higher burden on the healthcare system. Several interventions have taken place to improve chronic care management in general practices. Audit and feedback can promote change through the opportunity to reflect on performance. Practice accreditation, a specific form of an audit and feedback system, can help evaluate the quality of care provided by the practice and can be used to set up tailored improvement programs. In the Netherlands, primary care practices can take part in the Accreditation program developed by the Dutch College of General Practitioners. In this thesis, we focus on the evaluation of part of the Accreditation program that regards chronic care management. We describe several measurement aspects of the instrument that is used in the Accreditation program. Furthermore, we examine the impact of practice accreditation on chronic care management. The research questions of this thesis are summarized in Table 1.

Table 1. Overview of research questions of studies in this thesis

Chapter	Research questions
Chapter 2	To which extent does the domain clinical care of the Visitation Instrument Accreditation (VIA) used in the Dutch Accreditation program meet each of four reliability and validity criteria? 1. Correlation between the sets of performance indicators 2. Internal consistency within each set of performance indicators 3. Precision of indicator scores 4. Precision of benchmarks
Chapter 3	What is the relationship between sample size and precision? Can we increase precision with a composite performance score? How many indicators are needed minimally to achieve a certain level of precision?
Chapter 4	What is the effect of the Dutch Accreditation program on performance scores regarding diabetes, COPD and cardiovascular disease?
Chapter 5	What are the opinions of family physicians regarding the Dutch Accreditation program?
Chapter 6	Which patient characteristics are associated with performance scores?
Chapter 7	Which organizational factors are determinants of intermediate patient outcomes in routine diabetes care?

To derive meaningful conclusions from performance scores, it is important that the validity and reliability of the instrument used to evaluate chronic care management are adequate. In *chapter 2*, we examine four reliability and validity issues of the instrument: the correlation between the indicator sets, the internal consistency of each set, the precision of the performance scores and the precision of the benchmarks. We performed an observational study based on medical record data from 82 practices that participated in the Dutch Accreditation program. We included data on diabetes (9 indicators), chronic obstructive pulmonary disease (COPD, 5 indicators), asthma (4 indicators), cardiovascular disease (CVD, 8 indicators), influenza vaccination (1 indicator), cervical cancer screening (1 indicator) and prescription of antibiotics (1 indicator). Results show that correlations between these 7 sets of indicators were weak, which indicates that the chronic care section of the instrument provides a relatively broad representation of care. Furthermore, internal consistency of the sets with more than 1 indicator was adequate, which suggests that each set measures a coherent underlying concept. Regarding the precision of the performance scores we found that in order to achieve a maximum confidence limit of 10% (the true value of the population score lies between +10 and -10 percentage points from the measured score) we should base the score on records from 96 patients. To decrease the confidence limit to 5%, the number of patient records needed to calculate a score increased to 384. A sample of 233 practices was sufficient to achieve a maximum confidence limit of 5% around each benchmark. Given that practices are mainly interested in those scores that show the largest deviation from the benchmark to assess in which areas they can improve their care, a margin of 10% around their own score and 5% around the benchmark score is acceptable. However, when the intention is to use performance scores for other purposes, for example to rank practices, precision criteria become more stringent and current sample sizes are not sufficient.

In *chapter 3* we examine a possible solution for sample size issues with individual performance scores. We evaluated precision, which is expressed as the closeness of the estimated score to the 'true' population value. Typically, larger precision is achieved by including a larger sample of patients on which scores are based. However, there is a limit to the sample that can be included in routine clinical usage, based on time restrictions and prevalence of the condition measured. We examined the precision of individual performance scores and compared the results with precision of a composite performance score. We also looked at how many single performance scores should be included into the composite minimally to achieve an acceptable level of precision. We performed a descriptive statistical study using medical record data from 455 Dutch general practices that had taken part in the Dutch Accreditation program. We included indicators regarding diabetes, CVD and COPD and calculated performance scores per indicator and composite performance scores per topic, all

on practice level (the percentage of patients that met the indicator divided by the number of patients eligible for the indicator). Results show that for individual performance scores, a sample of 96 patients needs to be included to achieve a precision of 10 percentage points. When the scores are combined into a composite score, a sample of 43 (diabetes) to 89 (CVD) patients is sufficient to achieve a similar precision. The variation in required sample sizes between the different conditions was caused by the variability in scores within a condition, the closeness of the indicator mean scores to the maximum/minimum score of 100%/0% and the correlations between the performance scores included in the composite score. Furthermore, if it was accepted that a small percentage of practices (e.g. 10%) would not achieve the targeted level of precision, required sample sizes decreased further. We also found that a composite score derived from a relatively small set of 5 to 7 indicators could provide the same precision of measurement as a composite score made up from a larger number of indicators. We therefore conclude that the creation of a composite score based on a small number of indicators can provide sufficient precision. However, which level of precision is sufficient depends on the a priori reasons for measuring quality. Statistically, higher precision can be achieved by selecting indicators with lower inter-correlations and indictors that are closer to the maximum or minimum score level (100% or 0%). However, for assessment purposes it is important that scores do not approach a maximum level so that there is room for improvement in the indicator scores. Also, high inter-correlations of indicators are often favorable when one wants to achieve high validity of an indicator set. Careful consideration of measurement purposes and all aspects related to measurement and calculation of scores is advised. Precision needs to be greatest when there is an external summative assessment or when it is used in a payment mechanism. For formative quality improvement, such as the Dutch Accreditation program, current precision of single performance scores and precision of a composite score is sufficient.

In *chapter 4*, we present a comparative observational study in which we examine the impact of the Dutch Accreditation program on the quality of chronic care management. Although accreditation is widely used to assess and improve quality of healthcare providers, little is known about its effectiveness, particularly in primary care. The Dutch Accreditation program uses a 3-year cycle with extended measurements at the start of the first year. A first cohort of 69 practices was followed throughout the program with measurements at start and after three years (the start of the next cycle). A second cohort of 69 practices was measured at their start of the Accreditation program, during the same period of the post-assessment in the first cohort. Data were collected retrospectively over a one year period prior to the start. We evaluated the effect of practice accreditation in a design with two separate comparisons. In the first comparison we assessed improvements over time within the first cohort. In the

second comparison we assessed differences between the first and second cohort by comparing the follow-up measurement of the first cohort with the start measurement of the second cohort. We included measurements on diabetes, COPD and CVD. Results show that practices in the first cohort had improved on cholesterol measurement, feet examination and the prescription of lipid lowering medication (diabetes), spirometry measurement and the provision of a smoking cessation advice (COPD). As the follow-up measurement scores on these performance measures were similar to the baseline scores in the second cohort, it remains uncertain whether the effect can be attributed to the Accreditation program. Compared to the second cohort of newly starting practices, the practices in the first cohort performed better regarding achievement of a target cholesterol level (diabetes) and blood pressure level (CVD) and the registration of smoke status and measurement of glucose levels (CVD). These differences could provide evidence for the added value of the Accreditation program. The evidence for an added value of the Accreditation program was not as strong as expected by participants and stakeholders. This may be partially explained by other developments in the primary care field during our research period, such as the implementation of bundled payment. Both require data collection and offer feedback; measures used on diabetes, COPD and CVD are almost identical. However, we did find that general practices improved the quality of care for patients with the targeted chronic diseases. Continuous monitoring may have been beneficial for practices in order to maintain a high level of quality in some areas and improve further in others.

Chapter 5 presents a study on the opinions of family physicians regarding the Dutch Accreditation program. Evaluation of participant opinions can provide added insight in the mechanisms underlying the program and the impact of accreditation. A mixed methods design was used in which data from written questionnaires was combined with information gathered in interviews. All family physicians took part in the Dutch Accreditation program between 2006 and 2007 and also participated in a pay-for-performance study. The questionnaire contained open-ended and closed-ended questions. We used a framework approach in the interviews; the main themes and subthemes of the questionnaires and interviews were selected in advance, but some sub-themes were inductively derived during the analysis. The themes included motives for participation in the program, data collection, content of the measurement instrument, writing practice improvement plans, support and what the practice gained from participating. Results showed that family physicians mainly participated because they had a special interest in quality improvement and wanted to gain insight in their performance. While they were positive about the insight they gained into their practice management and clinical care, they found that the collection of data on medical practice was too time-consuming. Because of the lack of uniform registration, extraction from

electronic medical records was difficult. Regarding the content of the measurement instrument they were generally quite positive. However, they mentioned that the assessment instruments should put less emphasis on details and more on the underlying processes. Furthermore, physicians indicated that they needed more support in preparing a practice improvement plan. Overall, we found that family physicians are willing to take part in an accreditation program, but the process should not be too time-consuming and the instruments used should lend themselves to improving quality now and in the future. There needs to be a balance between measurability and perceived quality. Some physicians considered the drawbacks of the process to outweigh its advantages.

Patient characteristics may influence the performance scores measured in the Accreditation program. *Chapter 6* focuses on the relationship between patient characteristics and measurement of processes and outcomes of diabetes care in general practices. Patient age, gender, ethnicity and other socio-demographic factors are associated with diabetes outcomes. Furthermore, differences in the severity of the condition, defined as multi-morbidity, can also affect diabetes control. In order to compare performance measures between practices, adjustment or stratification is suggested for these factors. The strength of the association between multi-morbidity factors, other case mix factors and measures of processes and outcomes of care is not clear. Therefore, we investigated the associations of several patient factors (gender, age, body mass index (BMI), whether patient lives in a deprived area) and three multi-morbid conditions (cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD) and depression/anxiety) with diabetes care processes and outcomes. We included a total of 6 measures on whether HbA1c, systolic blood pressure and LDL cholesterol were measured and whether the measurement was within the target level described in the guidelines. We found that the models on processes of care showed moderate effects, with pooled areas under the curve (AUCs) that varied between 0.66 and 0.76. This suggests that patient factors had some impact on diabetes care indicators. Measurement of the number of diabetes consultations, as a measure of patient health service utilization, was the main contributor to the effects found (odds ratios between 5.6 and 14.5). Effects on outcome measures were small (AUCs varying from 0.57 to 0.61), which indicates that the model did not perform much better than chance. Our results do not support the need for case-mix correction or stratification suggested in previous literature. Therefore, we conclude that accounting for a certain margin of error around the scores without further correction may be more beneficial for users in practice.

The impact of accreditation on outcomes of care may vary between practices due to variations in organization that can be related to improved care. In *chapter 7* we examine

which organizational determinants are associated with improved diabetes care outcomes. Organizational determinants related to high-quality diabetes care, derived from a review of trials, were linked to specific items from validated questionnaires that were used in the Dutch Accreditation program. We included data from 362 practices that participated in the program between 2006 and 2009. We composed twelve measures and a composite score and examined the association of these measures with three diabetes outcomes: HbA1c, total cholesterol and systolic blood pressure. Results show that most organizational determinants were broadly implemented in our study population, indicating favorable conditions for high-quality diabetes care. A higher score on the composite measure of healthcare organization was associated with better systolic blood pressure control. Exploration of specific organizational factors identified only 3 significant effects in 36 combinations. These results from daily practice suggest that implementation of a combination of organizational determinants (rather than individual organizational items) is crucial for high-quality diabetes care.

Chapter 8 provides a general discussion of this thesis. All studies were based on observational data, which can be used to complement evidence from trials to enhance the generalizability. However, there were also several limitations to the use of retrospectively collected observational data, including the existence of missing values and possible selection bias. Also, practices already scored relatively high on many indicators of chronic care management, which may have limited the possibility of finding significant effects. In this chapter, we address three areas of discussion: (I) performance scores, (II) audit and feedback and accreditation, and (III) chronic care management. Regarding the performance scores we discuss whether they should be corrected or stratified for patient and practice characteristics. On the one hand it will increase face validity; on the other hand the associations that we found were small. It is debatable whether, for an individual practice, the benefit of correction outweighs the disadvantages (complex calculation methods and difficult interpretation of scores). We also suggest that Bayesian methods of analysis may provide interesting additional information to evaluate performance measures. These methods offer a solution to align statistical issues of assessment with issues of clinical relevance for the different stakeholders.

In the section on audit and feedback and accreditation, we discuss that whereas we found limited effects of accreditation on the processes and outcomes of chronic care management, the Accreditation program can also be of value to the general practice by means of creating more transparency. However, the costs and efforts to collect data should be in balance with the benefits. Since there is an increasing request to disclose quality data for the general public, we discuss the effects of public disclosure and pay-for-performance reported in other

studies: they may increase effects on outcomes, but there can also be more unintended consequences such as gaming and a lack of improvement in non-incentivized indicators. Also, there are several issues with validity and precision, since the requirements become more stringent for external summative assessment. To increase the chances of finding effects, the Dutch Accreditation program should be continuously tailored; data collection should be facilitated and existing barriers and facilitators to change should be addressed. The Dutch Accreditation program has been under development since the studies in this thesis were performed to address several of these issues.

Regarding chronic care management we discuss the use of guidelines. Performance scores that are used in the Dutch Accreditation program are derived from these guidelines. However, it is not desirable to strive for 100% scores. Not all guidelines are based on equally strong evidence, and in individual cases, deviation from the guideline may be beneficial for the patient. With the increasing focus on personalized medicine, deviations from the guideline may become more frequent. Also, guidelines are adjusted as the evidence is growing. This requires that physicians are flexible and willing to adjust their patient management. It is sensible to use performance measurements purely as an indication.

Samenvatting

‘Een goede statisticus gebruikt zijn analyses om te weten wanneer het volstaat deze achterwege te laten.’

Michel Wensing

In veel landen wordt praktijkaccreditatie aangeboden als strategie om de organisatie en uitvoering van zorg te verbeteren. Over de effectiviteit van praktijkaccreditatie is echter weinig bekend. In Nederland is in 2005 een accreditatieprogramma voor de huisartsenzorg gestart, de NHG-Praktijkaccreditering. In dit proefschrift onderzoeken we de impact van deze praktijkaccreditatie op de zorg voor chronische patiënten in de huisartspraktijk.

Hoofdstuk 1 bevat een algemene introductie over de kwaliteit van zorg voor chronische patiënten in de huisartspraktijk. Chronische zorg vormt een steeds groter aandeel van alle zorg geleverd door de huisartspraktijk. Er wordt verwacht dat zowel de prevalentie van de chronische aandoeningen, alsmede het aantal mensen met meerdere chronische aandoeningen, toe zal nemen in de komende jaren. Een belangrijke oorzaak hiervoor is de vergrijzing in de populatie. Dit brengt een extra druk op het zorggebruik met zich mee. Huisartspraktijken kunnen worden ondersteund bij het behouden van goede kwaliteit van chronische zorg of het verder verbeteren ervan met behulp van verschillende methoden. Zo kan audit en feedback bijdragen aan verbeteringen doordat het de mogelijkheid biedt het eigen handelen te evalueren. De NHG-Praktijkaccreditering maakt ook gebruik van een vorm van audit en feedback, ingebed in een 3-jarige cyclus waarin ook verbeterplannen ontwikkeld worden op basis van de audit en feedback.

Het is belangrijk dat de validiteit en betrouwbaarheid van de instrumenten, die gebruikt worden om de chronische zorg te evalueren, afdoende zijn om de scores juist te kunnen interpreteren en er zinvolle conclusies aan te kunnen verbinden. De meetinstrumenten die gebruikt worden binnen de NHG-Praktijkaccreditering om de kwaliteit van de chronische zorg in kaart te brengen, zijn in eerder onderzoek al gevalideerd op inhoud en klinische relevantie. Andere aspecten van betrouwbaarheid en validiteit zijn echter nog niet onderzocht.
In dit proefschrift onderzoeken we de samenhang van de meetinstrumenten onderling, hun interne consistentie en de precisie van de scores op de indicatoren en benchmarks. Door ook te kijken naar de precisie van samengestelde indicatorscores diepen we dit onderwerp nog wat meer uit; hiermee krijgen we een beter beeld van de zeggingskracht van indicatoren gegeven de grootte van de dataset waarop deze indicatorscores gebaseerd zijn.
Als we de betrouwbaarheid en validiteit van de meetinstrumenten onderzocht hebben is de volgende stap het kijken naar de effecten van accreditatie. In een volgende studie gaan we daarom in op het effect van de NHG-Praktijkaccreditering op processen en uitkomsten van zorg voor patiënten met diabetes, hart- en vaatziekten en COPD. Om een completer beeld te krijgen van de effecten van accreditatie onderzoeken we ook de meningen van huisartsen over de NHG-Praktijkaccreditering.

Om feedback correct te interpreteren is het van belang dat we op de hoogte zijn van factoren die de scores mogelijk kunnen beïnvloeden. Daarom bestuderen we tot slot de samenhang van patiëntkenmerken en organisatorische factoren met scores op het medisch handelen. Tabel 1 bevat de onderzoeksvragen van dit proefschrift.

Tabel 1. Onderzoeksvragen van de studies in dit proefschrift

Hoofdstuk	Onderzoeksvragen
Hoofdstuk 2	In hoeverre voldoet het medisch handelen onderdeel van het visitatie instrument accreditatie (VIA) aan 4 criteria betreffende betrouwbaarheid en validiteit? 1. Correlatie tussen de verschillende indicatorsets 2. Interne consistentie binnen elke indicatorset 3. Betrouwbaarheid van de indicatorscores 4. Betrouwbaarheid van de benchmarks
Hoofdstuk 3	Wat is de relatie tussen de steekproefgrootte en de precisie? Kan het creëren van een samengestelde indicatorscore leiden tot een grotere precisie? Hoeveel indicatoren zijn er minimaal benodigd om een zekere mate van precisie te bereiken?
Hoofdstuk 4	Wat is het effect van de NHG-Praktijkaccreditering op het medisch handelen in de praktijk wat betreft diabetes, COPD en hart- en vaatziekten?
Hoofdstuk 5	Wat is de opinie van de deelnemende huisartsen over de NHG-Praktijkaccreditering?
Hoofdstuk 6	Welke patiëntkenmerken zijn geassocieerd met medisch handelen scores?
Hoofdstuk 7	Welke organisatorische factoren hangen samen met medisch handelen scores op het gebied van diabetes?

In *hoofdstuk 2* beschrijven we een studie waarin we vier verschillende criteria over de betrouwbaarheid en validiteit van het instrument hebben onderzocht: de correlaties tussen verschillende indicatorsets, de interne consistentie van elke set, de betrouwbaarheid van de scores en de betrouwbaarheid van de benchmarks waarmee de scores vergeleken worden. We voerden een observationele studie uit die gebaseerd was op data uit het elektronisch patiënten dossier (EPD) van 82 praktijken die deelnamen aan de NHG-Praktijkaccreditering. De dataverzameling betrof diabetes (9 indicatoren), COPD (5 indicatoren), astma (4 indicatoren), hart- en vaatziekten (8 indicatoren), griepvaccinatie (1 indicator), baarmoederhalskanker screening (1 indicator) en antibiotica prescriptie (1 indicator). De resultaten tonen aan dat de correlaties tussen deze 7 indicatorsets laag waren, wat suggereert dat het onderdeel medisch handelen van het VIA-instrument een relatief breed beeld schetst waarin verschillende aspecten gerepresenteerd worden. Verder bleek dat de interne consistentie van de sets met meer dan 1 indicator adequaat was, wat een indicatie is dat elke set een coherent onderliggend concept meet. Wat betreft de betrouwbaarheid van de scores vonden we dat voor een maximale foutenmarge van 10% (dat wil zeggen dat de werkelijke onderliggende waarde ergens tussen de +10 en -10 punten van de gemeten score af ligt) de score gebaseerd moet zijn op gegevens van 96 patiënten. Om deze foutenmarge te verkleinen naar 5% steeg

het aantal benodigde patiënten naar 384 patiënten. Een steekproef van 233 praktijken was voldoende om benchmark scores te kunnen berekenen met een foutenmarge van maximaal 5%. Aangezien praktijken voornamelijk geïnteresseerd zijn in die scores waarin ze ver afwijken van de benchmark, om vast te stellen op welke punten ze zich het beste kunnen richten om de zorg te verbeteren, is een foutenmarge van 10% rond de eigen score en 5% rond de benchmark acceptabel. Wanneer de scores echter gebruikt worden voor andere doeleinden, bijvoorbeeld om een ranglijst te maken van de praktijkscores, dan worden de eisen die gesteld worden aan de scores wat betreft betrouwbaarheid strikter en is de huidige steekproefgrootte niet afdoende.

In *hoofdstuk 3* onderzochten we een mogelijke oplossing voor het relatief hoge aantal benodigde patiënten voor het berekenen van scores met een acceptabele foutenmarge. We evalueerden precisie, wat we definiëren als de mate waarin de geschatte score de onderliggende werkelijke populatie score weet te benaderen. Over het algemeen kan een betere precisie worden bereikt door een grotere steekproef te nemen om scores te berekenen. Echter, het aantal patiënten dat meegenomen kan worden in een steekproef wordt in de dagelijkse praktijk gelimiteerd door tijdsbelasting van dataverzameling, maar ook door de prevalentie van de betreffende aandoening. We onderzochten de precisie van losse indicatorscores en vergeleken de resultaten met de precisie van een samengestelde score waarin scores van meerdere indicatoren over één aandoening samengenomen werden. We onderzochten ook hoeveel scores op indicatoren er minimaal samengenomen moesten worden om een acceptabele mate van precisie te bereiken. We voerden een descriptieve statistische studie uit op basis van medische dossier data van 455 Nederlandse huisartspraktijken die deelnamen aan de NHG-Praktijkaccreditering. We namen indicatoren mee op het gebied van diabetes, hart- en vaatziekten en COPD; we berekenden losse scores per indicator en een samengestelde score per aandoening. Alle berekeningen werden gedaan op praktijkniveau (de indicatorscore betrof het percentage patiënten in de praktijk waarbij aan de indicator voldaan was). Uit de resultaten bleek dat voor losse scores gegevens van 96 patiënten benodigd waren om een precisie van 10% te behalen. Als de losse scores samengenomen werden in een score per aandoening bleek een steekproefgrootte van 43 (diabetes) tot 89 (hart- en vaatziekten) voldoende om een vergelijkbare precisie te bereiken. De verschillen in het benodigde aantal patiënten per aandoening werden veroorzaakt door I) de mate van spreiding tussen de scores binnen een aandoening, II) de mate waarin de gemiddelde indicatorscores in de buurt kwamen van de maximum/minimum score van 100%/0% en III) de correlaties tussen de losse scores die in de samengestelde score meegenomen werden. Verder vonden we dat het benodigde aantal patiënten om een score te berekenen nog verder daalt als we accepteren dat voor een klein deel van de praktijken

(bijvoorbeeld 10%) de nagestreefde mate van precisie niet bereikt wordt. Ook bleek dat met een samengestelde score die gebaseerd is op een relatief klein aantal indicatoren (5 tot 7) dezelfde mate van precisie bereikt kan worden als met een samengestelde score die gebaseerd is op een groter aantal indicatoren. We concluderen daarom dat met het berekenen van een samengestelde score gebaseerd op een klein aantal indicatoren een voldoende mate van precisie bereikt kan worden. Echter, welke mate van precisie gewenst is hangt af van de onderliggende doelen voor het meten van kwaliteit. Statistisch gezien kan een hogere precisie bereikt worden door indicatoren te selecteren die onderling minder sterk met elkaar samenhangen en waarvan de gemiddelde scores dichter bij het maximum of minimum (100%/0%) liggen. Anderzijds is het voor een meetinstrument juist belangrijk dat gemiddelde scores niet dicht bij het maximum liggen, zodat er voldoende ruimte is voor praktijken om te verbeteren. Daarnaast geeft men juist de voorkeur aan indicatoren die onderling sterker met elkaar samenhangen als het gaat om het samenstellen van een indicatorset met een hoge validiteit. Het is daarom van belang om bewuste keuzes te maken die gebaseerd zijn op het doel van de meting en op alle aspecten gerelateerd aan het meten en het berekenen van scores. Precisie moet vooral hoog zijn als de metingen gebruikt worden voor externe verantwoording of in een pay-for-performance systeem. Voor het in kaart brengen van de stand van zaken voor intern gebruik, zoals het geval is in de NHG-Praktijkaccreditering, is de huidige precisie van losse scores en samengestelde scores afdoende.

In *hoofdstuk 4* presenteren we een observationele cohortstudie waarin we de effecten van de NHG-Praktijkaccreditering op de kwaliteit van chronische zorg hebben onderzocht. Ondanks dat accreditatie wereldwijd wordt gebruikt om de kwaliteit van zorg vast te stellen en te verbeteren, is er vooral in de eerstelijnszorg weinig bekend over de effectiviteit ervan. De NHG-Praktijkaccreditering hanteert een 3-jarige cyclus met uitgebreide metingen aan het begin van het 1e jaar. Een eerste cohort van 69 praktijken werd gevolgd gedurende het programma, met metingen aan het begin en na 3 jaar (de start van de volgende cyclus). Een tweede cohort van 69 praktijken werd gemeten tijdens hun start van deelname aan de NHG-Praktijkaccreditering, in dezelfde periode als de nameting van het eerste cohort. Data werden retrospectief verzameld over het jaar voorafgaand aan de start. We evalueerden het effect van praktijkaccreditatie in een design met twee losse analyses. In de eerste analyse keken we naar verbeteringen in de tijd in het eerste cohort. In de tweede analyse keken we naar verschillen tussen het eerste en tweede cohort door de nameting van het eerste cohort te vergelijken met de startmeting van het tweede cohort. We namen metingen mee betreffende diabetes, COPD en hart- en vaatziekten. Uit de resultaten blijkt dat praktijken in het eerste cohort waren verbeterd wat betreft het meten van cholesterol, voetonderzoek en voorschrijven van lipide

verlagende medicamenten (diabetes), uitvoeren van een spirometrie (COPD) en het geven van advies om te stoppen met roken (COPD). Aangezien de nameting van het eerste cohort op deze indicatoren overeen kwam met de startmeting van het tweede cohort, is het onduidelijk of deze effecten toegewezen kunnen worden aan de NHG-Praktijkaccreditering. In vergelijking met het tweede cohort van nieuw startende praktijken, scoorden de praktijken in het eerste cohort beter op het bereiken van de cholesterol (diabetes) en bloeddruk (hart- en vaatziekten) streefwaarden; ook scoren ze beter op de registratie van rookgedrag en het meten van glucosewaarden (hart- en vaatziekten). Deze verschillen onderschrijven mogelijk de meerwaarde van de NHG-Praktijkaccreditering. Het bewijs voor de effectiviteit van de NHG-Praktijkaccreditering was niet zo sterk als van te voren verwacht door de deelnemers en andere betrokkenen. Mogelijk vormen de andere ontwikkelingen in de eerstelijnszorg gedurende de onderzoeksperiode, zoals de implementatie van zorggroepen, een gedeeltelijke verklaring voor deze bevindingen. Beide systemen vereisen dataverzameling en bieden feedback; de metingen betreffende diabetes, COPD en hart- en vaatziekten zijn bijna identiek. Uit onze studie blijkt dat de zorg voor chronische patiënten met de betreffende aandoeningen in de huisartspraktijk verbeterd is. Continue monitoring van zorg kan hebben bijgedragen aan praktijken om een goede kwaliteit van zorg te behouden ofwel verder te verbeteren in de verschillende aandachtsgebieden.

Hoofdstuk 5 beschrijft een studie over de ervaringen van huisartsen met de NHG-Praktijkaccreditering. Evaluatie van de opinie van deelnemers draagt bij aan het verkrijgen van inzicht in de onderliggende werkingsmechanismen van het programma en de effecten van accreditatie. We gebruikten een design met twee methoden; data van schriftelijke vragenlijsten werden gecombineerd met informatie uit interviews. Alle huisartsen namen deel aan de NHG-Praktijkaccreditering tussen 2006 en 2007 en namen ook deel in een pay-for-performance studie. De vragenlijst bevatte open en gesloten vragen. We gebruikten een zogenaamde ‘framework’ benadering in de interviews; de belangrijkste thema’s en subthema’s van de vragenlijsten en interviews waren van te voren vastgesteld, maar een aantal subthema’s werden inductief afgeleid uit de data tijdens de analyse. De thema’s die aan bod kwamen waren: redenen voor deelname aan het programma, data collectie, inhoud van de meetinstrumenten, het schrijven van verbeterplannen, ondersteuning en behaalde winst voor de praktijk. De resultaten tonen aan dat de huisartsen vooral deelnamen aan de accreditatie omdat ze een interesse hadden in kwaliteitsverbetering en omdat ze inzicht wilden krijgen in hun handelen. Alhoewel ze positief waren over het verkregen inzicht in hun praktijkorganisatie en medisch handelen, vonden ze dat de dataverzameling over het medisch handelen te veel tijd kostte. Aangezien er niet uniform geregistreerd was, was extractie van de data uit de elektronische dossiers lastig. Wat betreft de inhoud van de

meetinstrumenten was men over het algemeen positief. Wel gaf men aan dat de nadruk in de meetinstrumenten niet zozeer zou moeten liggen op de details, maar meer op de onderliggende processen. Verder gaven huisartsen aan dat ze meer ondersteuning zouden willen krijgen bij het schrijven van de verbeterplannen. In het algemeen vonden we dat huisartsen bereid zijn om deel te nemen aan een accreditatie, maar het proces moet niet al te tijdrovend zijn en de instrumenten moeten goed geschikt zijn om kwaliteit van zorg te verbeteren en te borgen voor de toekomst. Er moet een goede balans gevonden worden tussen meetbaarheid en ervaren kwaliteit. Sommige huisartsen vonden dat de voordelen van accreditatie niet opwogen tegen de nadelen van het proces.

Patiëntkenmerken kunnen mogelijk de scores op indicatoren gemeten in een accreditatieprogramma beïnvloeden. *Hoofdstuk 6* richt zich op de relatie tussen patiëntkenmerken en het meten van processen en uitkomsten van diabeteszorg in de huisartspraktijk. Leeftijd, geslacht, etniciteit en andere sociaal-demografische kenmerken van de patiënt zijn geassocieerd met uitkomsten van diabeteszorg. Daarnaast kunnen ook verschillen in de ernst van de aandoening, gedefinieerd als multi-morbiditeit, de mate van controle op diabetes uitkomstmaten beïnvloeden. Om scores tussen praktijken goed te kunnen vergelijken met elkaar wordt gesuggereerd dat scores gecorrigeerd of gestratificeerd moeten worden voor dergelijke patiëntfactoren. De mate van associatie tussen factoren van multi-morbiditeit, andere case-mix factoren en metingen van processen en uitkomsten van zorg is echter niet bekend. Daarom onderzochten we in deze studie de associaties van verscheidene patiëntkenmerken (leeftijd, geslacht, BMI, woonomgeving) en drie aandoeningen die vaak als co-morbiditeit bij diabetes voorkomen (hart- en vaatziekten, COPD en angst/depressie) met processen en uitkomsten van diabeteszorg. We namen in totaal 6 maten mee wat betreft of HbA1c, systolische bloeddruk en LDL cholesterol gemeten waren en of de HbA1c, systolische bloeddruk en LDL cholesterol waarden binnen de streefwaarden lagen zoals beschreven in de richtlijnen. In de modellen over de processen van zorg (het meten van de betreffende waarden) vonden we middelmatige effecten, met een oppervlak onder de curve (AUC) variërend van 0.66 tot 0.76. Dit suggereert dat patiëntkenmerken wel enige invloed hebben op de scores op diabetes procesindicatoren. De belangrijkste factor die bijdroeg aan het gevonden effect was het aantal diabetes gerelateerde consultaties, als indicatie van de mate waarin de patiënt gebruik maakt van de aangeboden zorg (odds ratio tussen de 5.6 en 14.5). Effecten op de uitkomstmaten waren klein (AUC variërend van 0.57 tot 0.61). Bij een dergelijke AUC presteert het model niet veel beter dan kans niveau. De resultaten in deze studie bieden geen ondersteuning voor de suggesties in eerdere literatuur dat case-mix correctie of stratificatie nodig is. We concluderen daarom dat het wellicht meer van nut is voor gebruikers in de praktijk om niet te corrigeren

voor dergelijke factoren maar wel rekening te houden met een zekere foutenmarge rondom de scores.

De effecten van praktijkaccreditatie op uitkomsten van zorg worden mogelijk beïnvloed door verschillen in praktijkorganisatie die gerelateerd kunnen zijn aan betere zorg. In *hoofdstuk 7* onderzoeken we welke organisatorische determinanten samenhangen met betere uitkomsten van diabeteszorg. Diverse organisatorische determinanten, die geassocieerd zijn met hoge kwaliteit van diabeteszorg en afgeleid zijn uit een review van trials, werden aan specifieke items van gevalideerde meetlijsten gekoppeld die gebruikt worden in de NHG-Praktijkaccreditering. We gebruikten data van 362 praktijken die deelnamen aan de NHG-Praktijkaccreditering tussen 2006 en 2009. We stelden 12 maten samen en berekenden een samengestelde score. Al deze maten werden onderzocht op hun associatie met 3 diabetes uitkomsten: HbA1c, totaal cholesterol en systolische bloeddrukwaarden. Uit de resultaten komt naar voren dat de meeste organisatorische determinanten breed geïmplementeerd waren in de deelnemende praktijken, wat suggereert dat de condities voor het leveren van hoge kwaliteit van diabeteszorg goed zijn. Een hogere score op de samengestelde maat van organisatorische determinanten was geassocieerd met betere waarden van systolische bloeddruk. Verder onderzoek naar de specifieke organisatorische determinanten liet zien dat er slechts 3 significante effecten waren van de in totaal 36 combinaties. Deze resultaten die gebaseerd zijn op routinematige zorg suggereren dat implementatie van een combinatie van organisatorische determinanten (in plaats van bepaalde losse organisatorische factoren) cruciaal is voor het bewerkstellingen van hoge kwaliteit van diabeteszorg.

Hoofdstuk 8 bevat een algehele discussie over de studies in dit proefschrift. Ten eerste bespreken we verschillende aspecten rondom data verzameling en de methoden van de studies. Alle studies zijn gedaan met behulp van observationele data, die gebruikt kunnen worden om de bewijslast uit trials aan te vullen om zo de generaliseerbaarheid te vergroten. Er zijn echter ook een aantal beperkingen aan het gebruik van retrospectief verzamelde observationele data, waaronder het ontbreken van waarden en mogelijke selectiebias. Verder scoorden praktijken al relatief hoog op veel van de gemeten indicatoren van chronische zorg, waardoor de mogelijkheden om significante effecten te vinden wellicht wat ingeperkt werden. In de discussie komen verder drie verschillende aandachtsgebieden aan bod: (I) scores op indicatoren, (II) audit en feedback in accreditatieprogramma's en (III) management van chronische zorg. Wat betreft de scores op indicatoren bediscussiëren we of we er goed aan doen de scores te corrigeren of stratificeren voor patiënt- en praktijkkenmerken. Aan de ene kant kan dit bijdragen aan een betere validiteit; aan de andere kant vonden we in onze studies slechts kleine associaties. Het is de vraag of voor een individuele praktijk de voordelen

van het corrigeren van scores opwegen tegen de nadelen (complexe rekenmethoden, lastige interpretatie van de scores). Verder suggereren we dat verdere analyses met Bayesiaanse statistiek mogelijk een interessante aanvulling kunnen vormen om scores op indicatoren te evalueren. Deze methode biedt een oplossing om in de statistiek en het meten ook verschillen in klinische relevantie voor diverse betrokken partijen mee te kunnen nemen.

In accreditatieprogramma's nemen audit en feedback een belangrijke plaats in. We bespreken dat de heterogene effecten op proces- en uitkomstmaten van chronische zorg die we in onze studies gevonden hebben overeenkomen met andere studies naar effecten van audit en feedback. Alhoewel accreditatie een geschikte methode kan zijn om kwaliteit te verbeteren, moet de moeite die het kost om data te verzamelen wel in evenwicht zijn met de voordelen van deelname aan accreditatie. Om bij te dragen aan het verhogen van de effecten van accreditatie, moeten we de NHG-Praktijkaccreditering continu aanpassen, het verzamelen van data moet verder gefaciliteerd worden en we moeten meer aandacht besteden aan bestaande belemmerende en bevorderende factoren om te veranderen. De NHG-Praktijkaccreditering is sinds de uitvoer van de studies in dit proefschrift ook continu in ontwikkeling om op de genoemde punten verder te verbeteren.
Aangezien er een toenemende vraag is naar het openbaar maken van informatie over kwaliteit van zorg, bespreken we dat het publiek ontsluiten van gegevens of het koppelen van een financiële bonus aan kwaliteitsdata (pay-for-performance) mogelijk de effecten op de uitkomst kan vergroten. Er kunnen echter ook meerdere onbedoelde gevolgen zijn, zoals 'gaming' en een gebrek aan verbeteringen op gebieden waar geen beloning aan gekoppeld is. Ook zijn er verschillende beperkingen op het gebied van validiteit en precisie, aangezien de eisen die aan de meetinstrumenten gesteld worden een stuk strikter worden bij gebruik voor externe verantwoording.

Wat betreft management van chronische zorg bespreken we de relatie tussen de richtlijnen en de gemeten scores op indicatoren, de interpretatie van de scores en of de meegenomen indicatoren een goede afspiegeling vormen voor de gehele chronische zorg. De indicatoren die gebruikt worden in de NHG-Praktijkaccreditering zijn afgeleid van richtlijnen; echter, het is niet wenselijk dat artsen gaan streven naar 100% behandelen volgens de richtlijnen. Er zijn verschillende redenen om af te wijken van de richtlijn. Niet alle richtlijnen zijn gebaseerd op even sterk wetenschappelijk bewijs en in individuele gevallen kan het afwijken van de richtlijn baat hebben voor de patiënt. Bovendien worden de richtlijnen continu aangepast naarmate de bewijslast groeit en verandert. Op elk moment in de tijd kan het meest recente bewijs mogelijk nog niet geïmplementeerd zijn in de richtlijnen. Dit vereist dat huisartsen zich flexibel opstellen en bereid zijn om hun handelen aan te passen. Het is ook van belang om de

instrumenten die gebruikt worden in een accreditatieprogramma continu aan te passen. Het kan een uitdaging zijn om scores juist te interpreteren; het is verstandig om de scores op indicatoren puur te zien als een indicatie. Wellicht is het nodig om regelmatig de beperkingen van indicatoren en hun scores uit te leggen aan de relevante betrokken partijen. In de toekomst kunnen we het meten van chronische zorg aanvullen door ook rekening te houden met aspecten zoals tijdigheid en door het toevoegen van indicatoren over geïndividualiseerde zorg en uitkomsten van kwaliteit van leven.

We concluderen dat het accreditatieprogramma een geschikt instrument kan zijn om kwaliteit te verbeteren. Het is van belang om te streven naar een optimale balans tussen het creëren van accurate en betrouwbare indicatorscores versus een acceptabele werkbelasting om de data te verzamelen. Het gebruik van het huidige programma voor interne evaluatie wordt aangeraden; na verdere evaluatie zou het programma in zijn geheel mogelijk ook geschikt kunnen zijn voor externe verantwoording.

Dankwoord

In Nieuw-Zeeland is er een gezegde dat wordt gebruikt wanneer een moeilijke taak gedaan moet worden: 'Show them that kiwi's can fly!'

Wikipedia

De afgelopen jaren heb ik me telkens geconcentreerd op de berg werk die nog voor me lag; nu ligt deze achter me en is het tijd om terug te kijken. Zonder de hulp van mijn promotieteam, collega's, familie en vrienden was dit proefschrift er niet geweest. Ik wil jullie dan ook graag bedanken.

Als eerste Jozé Braspenning. Ik vond het erg fijn dat je mij vanaf het begin af aan begeleid hebt, ik heb veel van je geleerd. In overlegmomenten bleek regelmatig dat we inhoudelijk wel op één lijn zaten, maar dat we nogal konden verschillen in onze benadering. Waar het voor mij logischer was om eerst de beren op de weg af te handelen om uiteindelijk tot een conclusie te komen, heb jij me geleerd meer te denken vanuit de kern en vanuit de boodschap die je over wilt brengen. Ook als mijn perfectionisme me weer eens in de weg zat bij het schrijven wist me je goed verder te helpen. Ik heb veel gehad aan de gesprekjes die we onderweg in de trein hadden; hartelijk dank voor de goede begeleiding.
Ook Michel Wensing wil ik hartelijk danken, ik had me geen betere promotor kunnen wensen. Je bijdrage in de overlegmomenten was altijd to-the-point en nuttig: met je helikopterview wist je goed aan te geven welke lijn we het beste zouden kunnen volgen. Ook in het schrijfproces heb ik veel aan je gehad. Mijn tweede co-promotor was René Wolters. Toen ik je vroeg of je in mijn promotieteam zou willen zitten twijfelde je; ik ben erg blij dat je toegestemd hebt. Niet alleen vanwege je achtergrond als huisarts (de enige in mijn promotieteam in een onderzoek waarin de huisartsenzorg centraal stond) maar ook als onderzoeker met een eigen mening heb je veel bijgedragen; ik heb dat erg gewaardeerd. Ook dank ik jullie alle drie voor de prettige samenwerking; jullie vulden elkaar onderling goed aan en vormden een fijn team.

Dit proefschrift had niet tot stand kunnen komen zonder de medewerking en financiële steun van het NHG, dank voor de fijne samenwerking en support. Ook de mensen van het NPA wil ik hartelijk danken voor hun medewerking om de data voor deze studie aan te leveren. Het was nogal eens puzzelen hoe we de data ook bruikbaar konden maken voor de wetenschap; met name Adri aan het Rot wil ik hartelijk danken voor alle uurtjes werk en de altijd weer verhelderende telefoontjes.

The first international article I wrote was the article on the precision of composite scores. We took quite some time to fine-tune the message of this article and perform all necessary calculations, but we got there in the end. David Reeves, I would like to thank you for your invaluable help in the analyses; thank you for all your time and efforts, also in the writing process. I still remember the various phone-calls and our meeting in Manchester. It was really stimulating to meet in person and work on the article together. You were very kind to

introduce Remco Feskens and myself to the University of Manchester, your colleagues and the local Indian cuisine. Also, I would like to thank Stephen Campbell. Your knowledge of developments in science and practice was crucial in choosing which direction to take with this article. I really appreciated your important contribution during the whole process. It was an honor and a great pleasure to work with you both.

Ook alle anderen die hebben meegeschreven aan de artikelen in dit proefschrift wil ik graag bedanken. Margriet Bouma, hartelijk dank voor je rol in het schrijven van de artikelen, maar ook als aanspreekpunt binnen het NHG. Mede dankzij jouw inzet heb ik al die jaren onderzoek kunnen doen voor het NHG en heeft het kunnen uitmonden in een promotie, dat waardeer ik enorm. Het contact was altijd erg prettig; de overlegmomenten lichtelijk chaotisch en met veel te veel te bespreken in korte tijd, maar wel heel leerzaam. Ik heb veel gehad aan het NHG als directe link met beleidsvorming voor de praktijk. Pieter van den Hombergh, het was erg leuk om jouw promotiewerk verder uit te kunnen bouwen en daar ook zo af en toe met jou over te kunnen sparren. Je enthousiasme werkte absoluut aanstekelijk. Al moest ik er wel even aan wennen dat ik ook op zondagochtend gebeld kon worden voor een inhoudelijke discussie… Gelukkig gebeurde dat niet te vaak. Dank voor de fijne samenwerking!
Remco Feskens, naast een waardevolle co-auteur was je ook een gezellige reisgenoot en kamergenoot, dank daarvoor. Het was erg handig iemand met zoveel kennis over statistiek naast me te hebben zitten toen ik net tot over mijn oren in een statistisch artikel zat; ook je ervaring met schrijven heeft me goed op weg geholpen. Wim de Grauw, dank voor je bijdrage aan het artikel over de organisatorische determinanten in de diabeteszorg. Je duidelijke visie leverde interessante discussies op bij het interpreteren van de gevonden resultaten. De samenwerking tussen IQ healthcare en de afdeling eerstelijnsgeneeskunde vond ik erg waardevol. Femke Atsma en Birgit Jansen, hartelijk dank voor jullie werk aan het artikel over de invloed van patiëntkenmerken op de diabeteszorg. Het heeft mij veel hoofdbrekens gescheeld om de data van LINH te kunnen gebruiken voor dit artikel met jullie hulp. Femke, dank voor het uitvoeren van de analyses en je deskundige input. Het was een plezier om met jullie samen te werken. Ook de overige co-auteurs, Kirsten Kirschner, Jako Burgers, Kees in ‘t Veld en Richard Grol wil ik graag bedanken voor de fijne samenwerking en hun waardevolle bijdrage aan de artikelen in dit proefschrift.

Ik wil graag alle medewerkers van IQ healthcare danken voor de fijne sfeer, leuke gesprekken en stimulerende vergaderingen. Ik ga niet iedereen bij naam noemen, maar een aantal mensen wil ik toch speciaal bedanken. Jolanda van Haren, ik heb altijd veel aan je gehad als aanspreekpunt op de afdeling, maar vooral in de laatste periode heb je me enorm geholpen met de praktische zaken rondom mijn proefschrift. Hartelijk dank voor je uitgebreide werk

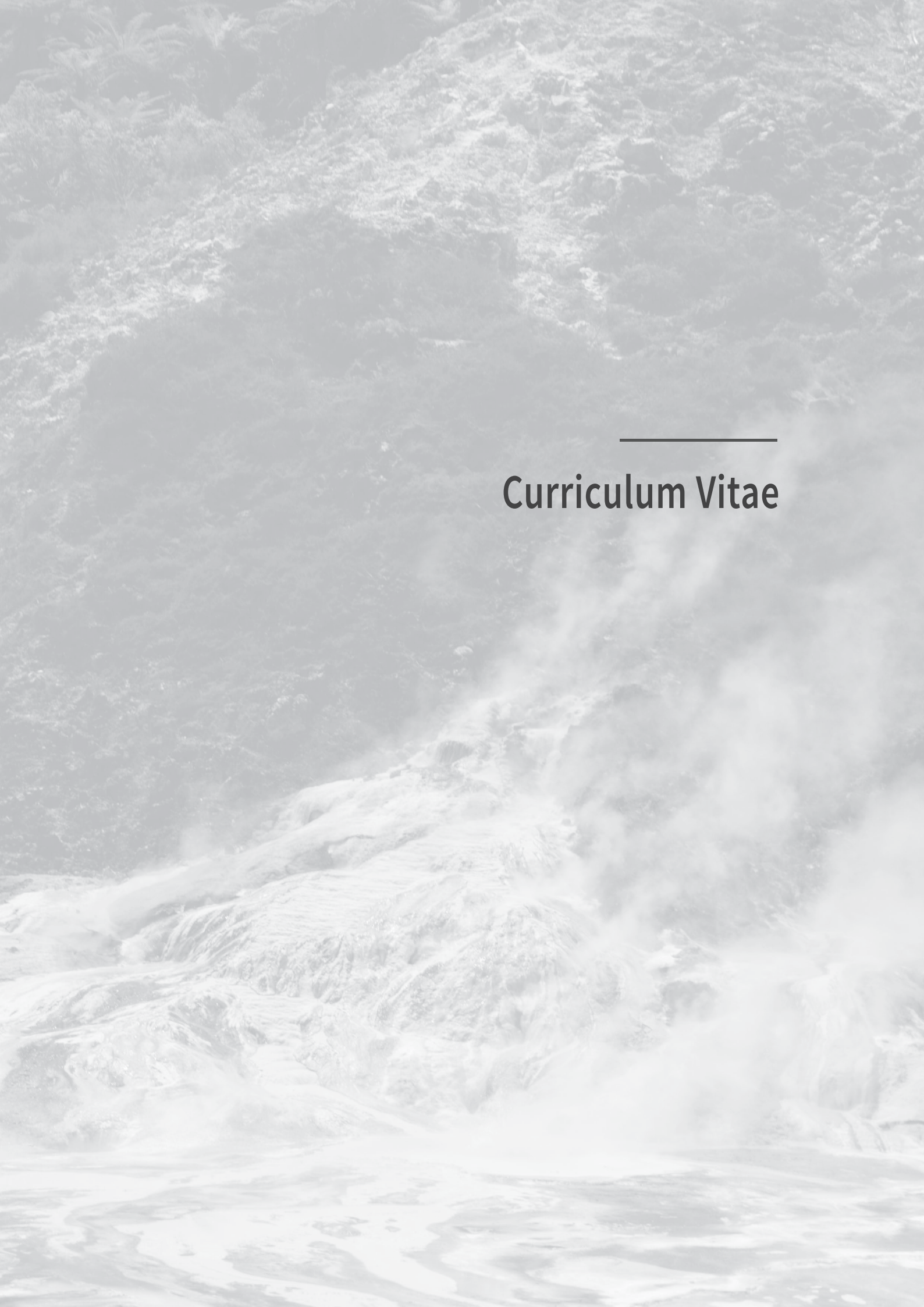
aan het vertalen van een hoofdstuk, het samenstellen van het proefschrift, de lay-out en de vele organisatorische regeldingetjes. Heel fijn!
Ik heb bij IQ ook een leuke tijd gehad in de werkgroep communicatie en kennisgroep indicatoren. Het was een plezier om met jullie samen te werken; ik vond het heel fijn op die manier ook wat voor de afdeling te kunnen betekenen. Ook naast het werk was IQ een leuke afdeling. Ik heb genoten van het zanggroepje tijdens de lunchpauzes. Lianne en Nicky, zingen was één van de overeenkomsten die we hadden, maar ook de stimulerende gesprekken met jullie heb ik altijd erg gewaardeerd. Het bowlen/paintballen/film kijken maakte ook dat je je collega's nog eens op een andere manier leerde kennen, dit was altijd erg gezellig.
Ook de bedrijfsuitjes waren doorgaans erg geslaagd, met als persoonlijk hoogtepunt het jaar waarin we 'Wie is de Mol' deden en ik de beste mol van de afdeling bleek te zijn; ik wist niet dat ik het in me had. Ik herinner me verder nog goed het weekendje skiën met Kirsten, Renate, Linda, Selma, Lianne en Lucy, dat was erg geslaagd, ik heb ervan genoten. Ook de ISQua en Wonca congressen in Genève, Parijs en Malaga waren een geweldige ervaring, dank aan alle collega's die deze reizen mede zo aangenaam gemaakt hebben.
Door de jaren heen heb ik verschillende kamergenoten gehad: Marije, Anouk, Gijs, Remco, Nancy, Wytske en Lydia. Jullie waren geweldige kamergenootjes om mee te sparren en lief en leed te delen. Hartelijk dank voor alles! Gijs en Wytske, super dat jullie mijn paranimfen wilden zijn! Aan alle andere collega's, te veel om iedereen bij naam te noemen: hartelijk dank allemaal voor de gezellige tijd die ik bij IQ gehad heb!

Marianne Gullberg, jou wil ik hier graag bedanken omdat je in zekere zin de aanstichtster van mijn wetenschappelijke carrière bent geweest toen ik nog als onderzoeksassistent bij het Max Planck Instituut werkte. Je grote enthousiasme voor onderzoek en enorme kennis werkte zeker aanstekelijk. Ik heb nu mijn 'driver's license' binnen!

Op donderdagavond is het altijd weer een plezier de frustraties van die week er even uit te slaan tijdens het volleyen. Ik waardeer jullie vriendschap enorm, de leuke gesprekjes in de kleedkamer en de gezelligheid tijdens het sporten. Ook op de Veerkracht heb ik leuke mensen om me heen, fijn dat we elkaar hebben leren kennen. Verder wil ik hier nog het oude clubje van manege Neerbosch noemen: super leuk dat we nog altijd zo nu en dan bij elkaar komen om weer bij te praten, ik hoop dat we dat nog lang volhouden.
Dieke, dank voor je fijne ondersteunende mailtjes vanuit Noorwegen, altijd weer fijn even van elkaar te horen! En tot slot Elodie: je bent een geweldige vriendin, ik waardeer je enorm. Het is erg fijn om samen goede gesprekken te hebben met een theetje op de bank; super dat je er zo voor me bent. We moeten nodig weer eens afspreken voor een spelletjesavond!

Lieve familie, dank voor alles. Ik stel ons prettige contact erg op prijs, dank voor jullie interesse! Het is heel fijn om te merken dat jullie er altijd voor ons en voor elkaar zijn. Lieve Jaap en Greet, pap en mam, ik waardeer het enorm dat jullie al mijn hele leven achter me staan in alles wat ik doe, er voor me zijn en van me houden zoals ik ben. Ik zou me geen betere ouders kunnen wensen, ik hou van jullie!

En tot slot mijn lieve man, vriend, steun en toeverlaat: Sander. Ik ben zo blij met jou! We hebben geen woorden nodig om elkaar te begrijpen, je voelt het altijd precies aan hoe het met me is. Het is heel fijn om alles met je te kunnen delen. Ik zie er naar uit de rest van ons leven samen door te brengen, ik hou van je!



Curriculum Vitae

Arna Linda van Doorn - Klomberg werd op 30 oktober 1981 geboren in Nijmegen. Ze groeide op in Beuningen bij haar ouders Jaap en Greet van Doorn. In 1999 behaalde zij haar VWO-diploma aan het Stedelijk Gymnasium in Nijmegen.

Tussen 1999 en 2005 studeerde ze psychologie aan de Radboud Universiteit Nijmegen. Ze koos voor de afstudeerrichting neuro- en revalidatiepsychologie en volgde diverse extra vakken op het gebied van onderzoek en klinische psychologie. Ook was ze actief in de studievereniging Homunculus als bestuurslid en penningmeester. Haar afstudeerscriptie richtte zich op problemen met werkwoordverbuigingen bij mensen met een afasie van Broca. Tijdens haar studie begon ze ook met werk als onderzoeksmedewerker bij het Max Planck Instituut; tot eind 2007 heeft ze daar gewerkt in de onderzoeksgroep meertaligheid.

In december 2007 begon ze haar werk als junior onderzoeker bij de afdeling IQ healthcare van het Radboudumc. Aanvankelijk verrichte zij projectwerk voor het Nederlands Huisartsen Genootschap; dit projectwerk mondde uit in het promotietraject waarvan dit proefschrift het resultaat is. Naast haar promotietraject was ze betrokken bij diverse andere projecten, waaronder het ontwikkelen van publieke indicatoren voor de huisartsenzorg op het gebied van het medisch handelen en het ontwikkelen van een vragenlijst rondom de organisatie van cervix screening in de huisartsenpraktijk. Binnen de afdeling zette ze zich in als lid van de communicatie werkgroep en de kennisgroep transparantie. Sinds 2014 werkt ze als docent bij de Hogeschool Arnhem Nijmegen, bij de opleiding Master Physician Assistant. Ze geeft hier onderwijs op het gebied van wetenschappelijk onderzoek. Ook begeleidt ze studenten bij hun afstudeeronderzoek.

Arna is sinds 2010 getrouwd met Sander, ze wonen in Nederasselt.